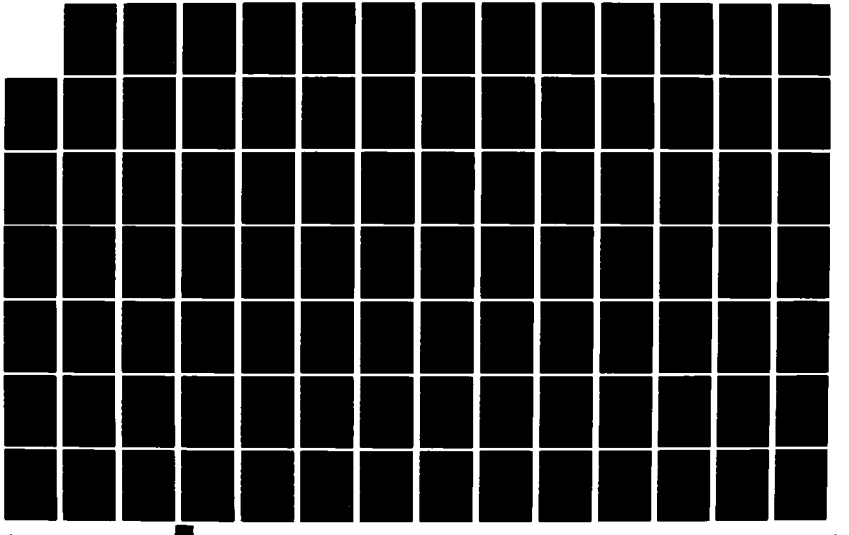
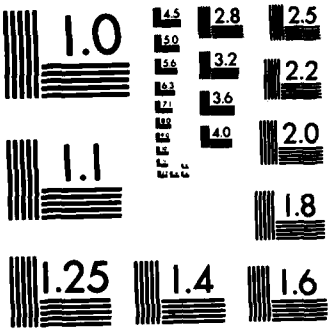
AD-A126 763 MASS SPECTROMETRIC RAPID DIAGNOSIS OF INFECTIOUS DISEASES(U) STATE UNIV OF NEW YORK AT BUFFALO SCHOOL OF MEDICINE M ANBAR ET AL. MAR 81 DAMD17-78-C-8035 1/2

UNCLASSIFIED .F/G 6/5 NL





FINAL REPORT

Michael Anbar
James McReynolds
Michael Scanlon
Robert Abbott
Lev Verkh

March, 1981

MASS SPECTROMETRIC RAPID DIAGNOSIS OF INFECTIOUS DISEASES

Supported by

U.S. ARMY MEDICAL RESEARCH AND DEVELOPMENT COMMAND
Fort Detrick, Maryland

Contract DAMD17-78-C-8035

School of Medicine
State University of New York at Buffalo
Buffalo, New York

UNCLASSIFIED

Department of Biophysical Sciences
School of Medicine
State University of New York at Buffalo
Buffalo, New York 14214

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY PRACTICABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

TABLE OF CONTENTS

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER

TITLE (and Subtitle)

Mass Spectrometric Rapid Diagnosis of Infectious Diseases

5. TYPE OF REPORT & PERIOD COVERED

Final Report Feb. 1 1980 to Feb. 28, 1981

6. PERFORMING ORG. REPORT NUMBER

AUTHOR(s)

Michael Anbar, Robert Abbott, James McReynolds, Lev Verkh, Michael Scanlon

8. CONTRACT OR GRANT NUMBER(S)

DAMD17-78-C-8035

PERFORMING ORGANIZATION NAME AND ADDRESS

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

CONTROLLING OFFICE NAME AND ADDRESS

U.S. Army Medical Research and Development Command, Fort Detrick, Maryland

12. REPORT DATE: March, 1981

13. NO. OF PAGES: 105

MONITORING AGENCY NAME & ADDRESS (if diff. from Controlling Office)

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited

DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

SUPPLEMENTARY NOTES

KEY WORDS (Continue on reverse side if necessary and identify by block number)

Clinical diagnosis, infectious diseases, metabolite profiles, urine, field ionization mass spectrometry, multi-component analysis, molecular weight profiles

ABSTRACT (Continue on reverse side if necessary and identify by block number)

Using new chemical procedures, mass spectrometric instrumentation and appropriate computerized data analysis, the diagnosis of a number of infectious diseases, through the molecular weight profile of neutral metabolites urine, was demonstrated. Longitudinal studies on human volunteers infected with sandfly fever showed the appearance of a characteristic pattern prior to the onset of clinical symptoms which persisted some time after the symptoms have subsided. Tissue cultures infected with polio virus exhibited a characteristic pattern demonstrable within hours after exposure.

DD FORM 1 JAN 73 1473

EXECUTIVE SUMMARY

This report describes the second and final phase of a research program aimed at the utilization of multicomponent analysis by non-fragmenting mass spectrometry for diagnosis of infectious diseases. During this phase it has been demonstrated that metabolic profiles of the host can be used to identify infected subjects, to differentiate between patients with different infections, to detect a viral infection prior to the onset of clinical symptoms, and to demonstrate the existence of a viral infection some time after the clinical symptoms have subsided. Further, it has been demonstrated that mass spectrometric multicomponent analysis can be a highly useful tool in the clinical laboratory by detecting the presence of a virus through its effect on the metabolism of tissue cultured cells. A diagnostic biochemical pattern seems to be distinguishable within a few hours, which is significantly faster than by the presently used morphological changes. Also, new sample preparation techniques, instrumentation, and computerized statistical analysis have been developed during this phase of the program.

I. INTRODUCTION

This is the third annual and final report on the second phase of a program on "Mass Spectrometric Rapid Diagnosis of Infectious Diseases" conducted under contract DAMD 177808035 with the U.S. Army Medical Research and Development Command. This report covers the period February 1, 1980 through February 28, 1981 and concludes the research effort under this contract.

The purpose of the program has been to develop a methodology for the rapid diagnosis of infectious diseases based on non-fragmenting mass spectrometry. The experimental approach is based on a novel methodology and instrumentation developed by the Principal Investigator at Standford Research Institute (SRI). This methodology has been shown to facilitate the detection of specific metabolic aberrations that occur in the host as a result of an infectious process. It might also be applied to the detection of viruses in tissue cultures and to the identification of microorganisms by their chemical constituents or by their characteristic metabolic products.

The capability of making a rapid and reliable diagnosis of infectious diseases at an early stage and at low cost would be of especially great value to the military where large numbers of army personnel are stationed in confined areas and their continuing health is crucial to carrying out their objectives. Early and reliable diagnosis of an infectious disease could prevent the spread of disease to larger groups of military and civilian personnel.

Viral infections are generally harder to diagnose than bacterial diseases and in many cases it is even difficult to establish whether a given set of pathophysiological clinical symptoms are associated with a viral infection. Further, the identification of presence of a virus and the characterization of

a detected virus has to be carried out in tissue cultures in assays that may take over 24 hours, or even 48 hours, to complete. The methodology developed in this program of research is directed (1) at establishing if a viral infection in humans lead to a specific host biochemical reaction expressed in a characteristic change in the metabolic profile of urine, and (2) at establishing the presence and possibly the identity of viruses in vitro by monitoring early changes in the metabolic profiles of tissue culture media. We have reason to expect the detection of characteristic biochemical changes in tissue culture media before any morphological changes can be detected.

During the second phase of this program, we have achieved a number of critical objectives. After further improvements and optimization of the sample preparation techniques and after finding optimal conditions for mass spectrometric analysis (utilizing a double focussing configuration), we have devoted a substantial effort to compare and critically analyze alternative statistical multivariate diagnostic procedures. Further, a substantial part of the effort has been devoted to the analysis of clinical samples and thus to the evaluation of the potential usefulness of our methodology as a clinical diagnostic technique. We have demonstrated a high degree of success in separating a number of different groups of patients by the metabolic profiles of their urines. These included patients with alcoholic liver disease, children with pneumonia and children with virus induced diarrhea; the patients could be readily differentiated by comparison with healthy subjects of the corresponding age groups with practically no false-positives or false-negatives. Next we have carried out a longitudinal study on urine samples obtained from groups of volunteer subjects vaccinated with live virus of sandfly fever and dengue fever, who were followed up for a number of weeks. This study suggests that a diagnostic pattern associated with the infection can be detected before the onset of clinical symptoms and it

subsides after their disappearance (Clin. Chem. 29, 1443 (1980)). Also, significant differences in the rate of individual reaction to the infection could be observed.

In another series of studies we have shown that the molecular weight profile of a human and animal tissue culture medium exhibits significant changes when the cells are infected with a virus, within hours following infection. Human lung tissue cultures infected with polio myelitis demonstrated a diagnostically useful pattern 6 hours following exposure to the virus (Clin. Chem. 26, 1443 (1980)). These feasibility studies should now be continued to establish the scope and limitations of this early detection technique.

In brief, during the second phase of this program, preliminary experiments have shown that metabolic profiles of the host can be used to identify infected subjects, to differentiate between patients with different infections, to detect a viral infection prior to the onset of clinical symptoms, and to demonstrate the existence of a viral infection some time after the clinical symptoms have subsided. Further, it has been demonstrated that mass spectometric multicomponent analysis can be a highly useful tool in the clinical laboratory by detecting the presence of a virus through its effect on the metabolism of tissue cultured cells. A diagnostic biochemical pattern seems to be distinguishable within a few hours, which is significantly faster than by the presently used morphological changes.

II. SUMMARY OF ACCOMPLISHMENTS

The objective of the proposed program of research has been to answer the following questions:

1. Do infectious diseases exhibit characteristic concentration patterns of metabolites in urine?
2. Can a rapid and reliable pattern recognition analytical technique, capable of providing such diagnostic information based on non-fragmenting mass spectrometry, be developed?
3. Can the technique make the diagnosis with sufficient sensitivity (small percentage of false negatives) while being highly specific (virtually nil false positives)?
4. To what extent are the changes in the chemical construction of these biological fluids indicative of the severity of viral infection?
5. To what degree can the chemical aberrations in urine be used as indicators for recovery, and to what degree can they be useful for the identification of post-infection carriers?
6. Can viruses be detected through biochemical changes induced in cultured cells?
7. Can viruses be identified through characteristic metabolites released from tissue cultured cells into controlled artificial media, within a few hours of incubation?
8. Can the chemical identity of the metabolites constituting a given diagnostic pattern be determined, to allow their determination by analytical techniques other than mass spectrometry.

The first three questions have been answered affirmatively during the previous phases of the program (Clin. Chem. 22, 1503 (1976)). The results obtained recently corroborate these conclusions in an unambiguous manner. Questions 4, 5 and 6 are still open. However, the preliminary results obtained by the longitudinal studies last year are very encouraging - demonstrating a diagnostically significant biochemical pattern in urine prior to the onset of clinical symptoms of a viral infection - a pattern that may persist for some time after the clinical symptoms have subsided. The preliminary answer to question 6 is also positive, although much more work is required to substantiate these findings. The answers to questions 7 and 8 are still pending and will be addressed in the proposed next phase of the program.

In the second phase of this program we proposed to accomplish the following:

1. Substantiate the findings on the rapid differential diagnosis of hepatic, pulmonary, and urinary infections.
2. Carry out longitudinal study on patients suffering from viral infections from the onset of the disease through stages of recovery and reconvalescence.
3. Improve the statistical data handling techniques to cope better and faster with the diagnostic problems, and, in particular, develop pattern recognition techniques capable of distinguishing the various biochemical factors detected in subjects with infectious diseases.

Each of these tests has been accomplished as described in part III of this report. We have substantiated the preliminary findings that characteristic metabolic profiles detectable by mass spectrometry are associated with various infections. We have shown the development in healthy individuals of a characteristic pattern following infection preceding the onset of clinical symptoms. Moreover, this characteristic metabolic profile seems to persist

through the stage of recovery (Clin. Chem. 26, 1443 (1980)). We have made significant progress in understanding and adapting alternative diagnostic statistical classification techniques. We have also tested the possibility of early detection of the presence of a virus in a tissue culture by the monitoring of the medium and the results are highly encouraging, so that we plan to devote to this problem a substantial part of the next phase's effort.

In the second phase of this program, we were interested primarily in establishing the feasibility of the diagnostic pattern recognition technique, both *in vivo* and *in vitro*. We did not, however, study in detail each of the parameters which determine the quality of the metabolic profiles in order to find optimal conditions of speed of analysis, including data handling, accuracy and precision, and intersample memory effects. This is true of all the stages of the analytical procedures from the stage of collection of the biological fluid, its short term storage, its long term storage, the pre-treatment of samples in preparation for mass spectrometry and then the mass spectrometry itself and its computerized data processing. Such an optimization will not only streamline and increase the throughput of the analytical laboratory but it will also further improve the sensitivity and specificity of the technique and make it applicable to monitor more subtle pathological changes.

III. TECHNICAL BACKGROUND

A. INTRODUCTION

During this second phase we have accomplished the following tasks:

1. Developed and tested sample sterilization and storage procedures.
2. Developed new, simpler sample preparation techniques, including one to handle tissue culture media.
3. Improved the mass spectrometric analysis procedure.
4. Studied and compared different ionization procedures and sample handling techniques.
5. Improved the interfacing with the INCOS-NOVA and the CYBER 173 computers.
6. Tested and evaluated different statistical analysis classification procedures.
7. Analysed samples of children and adults with different pathological problems and demonstrated the efficiency of our diagnostic procedure.
8. Analysed samples of series of samples obtained from virus infected volunteers over a period of 4 weeks, and demonstrated the appearance of a pathological pattern which disappeared after the disappearance of the clinical symptoms. Different sample pre-treatment procedures and ionization methods have been tested in this study.
9. Analysed media of human and animal tissue cultures infected with polio virus and demonstrated a diagnostically useful pattern within hours following exposure to the virus.

Results of tasks 8 and 9 have been published recently in Clinical Chemistry (Appendix A). However, this technical report includes more recent experimental results.

B. EXPERIMENTAL TECHNIQUES AND PROCEDURES

The following is an updated description of our experimental techniques and procedures. Although some of the material described below was included in the previous annual reports significant changes have been introduced since that time in a number of phases in the analytical procedures to warrant reporting anew the whole analytical procedure.

1. Sample Collection and Storage Procedure

Urine samples were collected by hospital staff in 12 ml plastic collection tubes with self-sealing caps (Kova-Tubes™) and kept frozen until analysis. As an enzyme denaturant, bacteriocide and potential virus inactivator, 0.1 ml of 0.5 M $ZnSO_4$ or $HgCl_2$ was placed in tubes given to the hospital or clinic. This results in a final concentration of approximately 0.005 M $ZnSO_4$ in the collected urine.

After reaching room temperature, an equimolar amount of EDTA is added to chelate the excess zinc or mercury ions, and the pH of the sample is adjusted to 7.2 with HCl or NaOH. Chelation of the excess zinc with EDTA, avoids precipitation of $Zn(OH)_2$ during pH adjustments. Such precipitation could potentially alter profile patterns by co-precipitating organic constituents in the urine.

Samples (5 ml aliquots) were thawed and 0.005M $ZnSO_4$ was added to samples that were collected without this preservative. The presence of zinc was found to have a minimal but detectable effect on the molecular weight profile, so all samples are presently processed with this reagent added. In the future $HgCl_2$ will be substituted for $ZnSO_4$ in this protocol to eliminate the possibility of a viable viral contamination (see below).

Dróplets of thawed urine are also tested using Ames Multistix^{R} reagent strips to assay pH, protein, blood, glucose, ketone bodies, bilirubin and urobilinogen. These results are recorded for correlation with hospital testing of the same sample, and to identify samples that could be excluded from data sets on the basis of abnormal kidney function.

In conjunction with the in vitro tissue culture experiments both $ZnSO_4$ and $HgCl_2$ were tested by Dr. Howard Faden of the Childrens Hospital for their ability to inactivate polio virus (Mahoney polio strain). The viral solution was supplemented with one or the other of the salts and then treated with EDTA to remove excess metal ions that were themselves cytotoxic. Viral solution treated with 5 mM $HgCl_2$ did not cause infection when added to a culture of human embryonic lung tissue. On the other hand, concentrations of up to 10 millimolar $ZnSO_4$ were not effective in inactivating this virus. The latter findings indicate the preference of $HgCl_2$ as a sterilizing additive.

A titration curve of urine indicated that pH values less than 2, near 7.2 and above 10 are likely to yield reproducible profile patterns. In intermediate pH regions the extraction yields of partially ionized constituents are highly pH dependent. In a separate series of experiments, using the Wilcoxon-WNI test to evaluate the results, it was determined that in urine extracts insignificant pattern changes occurred within a pH range of $\pm$ 0.5 pH unit of 7.2.

We have studied the effects of various means of sample storage. This study is outlined in Figure 1. A volume of urine was collected from a healthy adult male. This volume was divided into 5 ml aliquots that were alternatively refrigerated, frozen and left standing at room temperature without preservative, in capped, 12 ml plastic Kova Tubes^{TM}. Three samples were prepared immediately from the fresh urine. Samples stored under the

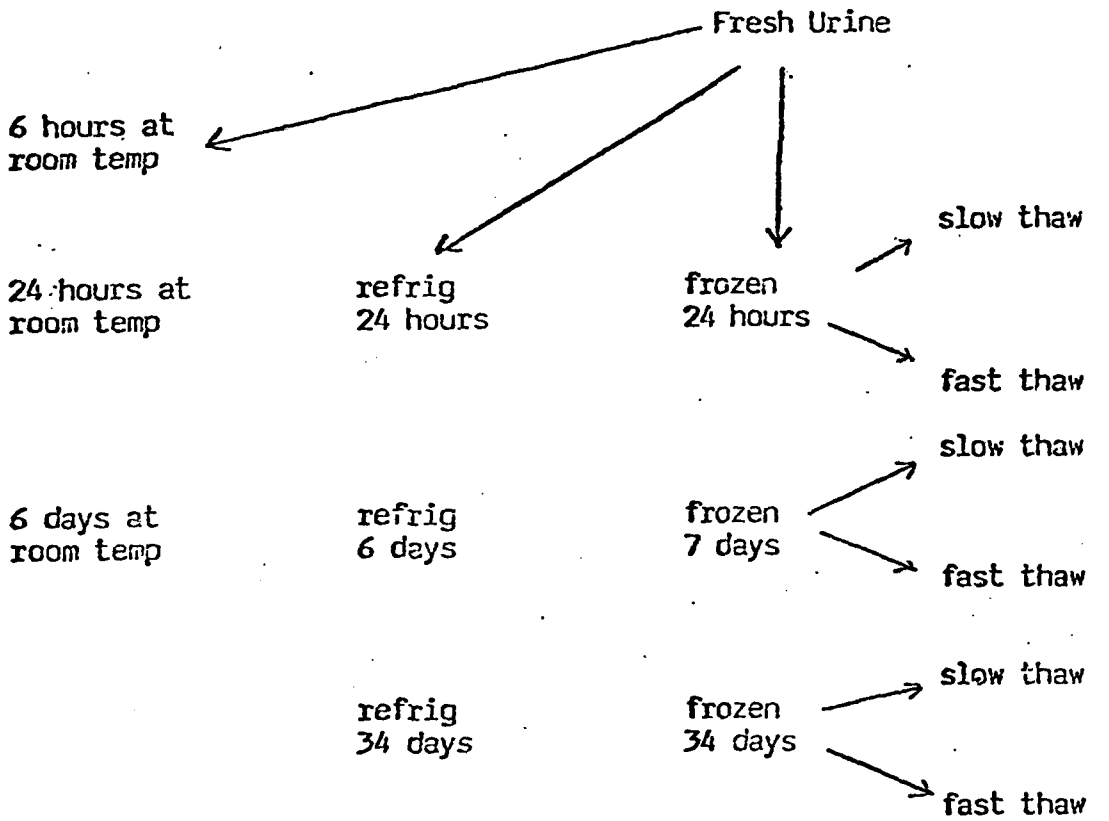
conditions outlined in Figure 1 (refrigerated at +5°C and -20°C respectively) were subsequently prepared in triplicate for mass spectrometric analysis.

Although in our short term study no significant changes were observed in the mass spectrometric profiles of samples refrigerated up to 6 days or frozen up to 34 days, later experience gained from urine sample stored at -20°C for 3 to 12 months indicate substantial changes. As consequence we have recently changed our storage procedures to use a deep freeze at -76°C.

Two different methods were employed to defrost the frozen samples. At each time indicated, three tubes were allowed to defrost by standing at room temperature for approximately 45 minutes. Three other tubes were more rapidly thawed by holding them under running water. No significant effect of the rate of thawing was detected.

FIGURE 1

URINE STORAGE STUDY



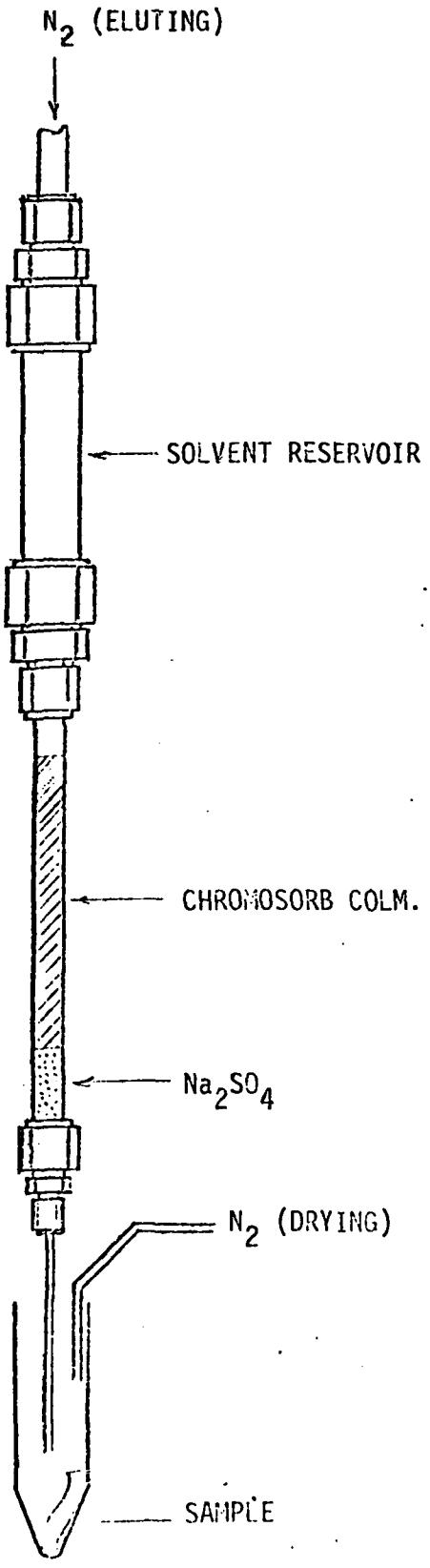
2. Sample Preparation Procedure

Samples are applied to the top of a 6 mm O.D. x 25 cm glass column containing 18 cm of prewashed Chromosorb P above 2.5 cm of Na_2SO_4 and a glass wool plug. The column outlet, a 10 cm length of 0.25 mm I.D. stainless steel capillary tubing, and a solvent reservoir of 12 cm O.D. x 100 cm glass tubing attached to the top of the column, are assembled using stainless steel swagelok™ unions and teflon™ ferrules (see Fig. 2). The sample is applied to the column using nitrogen pressure supplied through a swagelok™ fitting at the top of the column. The Na_2SO_4 serves to retain any water that may emerge from the chromosorb. After the sample is adsorbed, 5 mls of dichloromethane are added to the reservoir and forced through the Chromosorb column with N_2 pressure at a flow rate of 0.5 ml/min. The emerging eluate is continuously absorbed onto a 1 mm x 2 cm strip of Whatman GP/A glass fiber filter paper at the bottom of a conical tube. The eluate on the paper is continuously concentrated by a stream of dry N_2 directed toward the bottom of the collection tube. The dried eluate is stored in a glass vial until mass spectrometric analysis.

The filter paper replaces a micro column of chromosorb previously used (see 1979's report). The glass wool plugs of those sample columns frequently loosened and resulted in sample loss during loading into the mass spectrometer inlet probe. The glass fiber paper has a lower background than the chromosorb column and gives evaporation profiles similar to those obtained with the micro columns.

Samples for chemical ionization mass spectrometry were prepared by incubating 1 ml aliquots of urine with 0.5 ml of phosphate buffer (pH 6.8) and 0.5 ml urease solution (37I units/ml) for one hour at room temperature.

FIGURE 2



SAMPLE EXTRACTION APPARATUS

Following digestion of urea the pH was adjusted to pH 7 with 1N HCl. An aliquot of sample equivalent to 1 to 5 microliters of the undiluted urine was dried on a 2 mm x 1 cm strip of glass filter paper. Samples were allowed to dry in air at room temperature and were stored for periods of one day or less at room temperature until analysis.

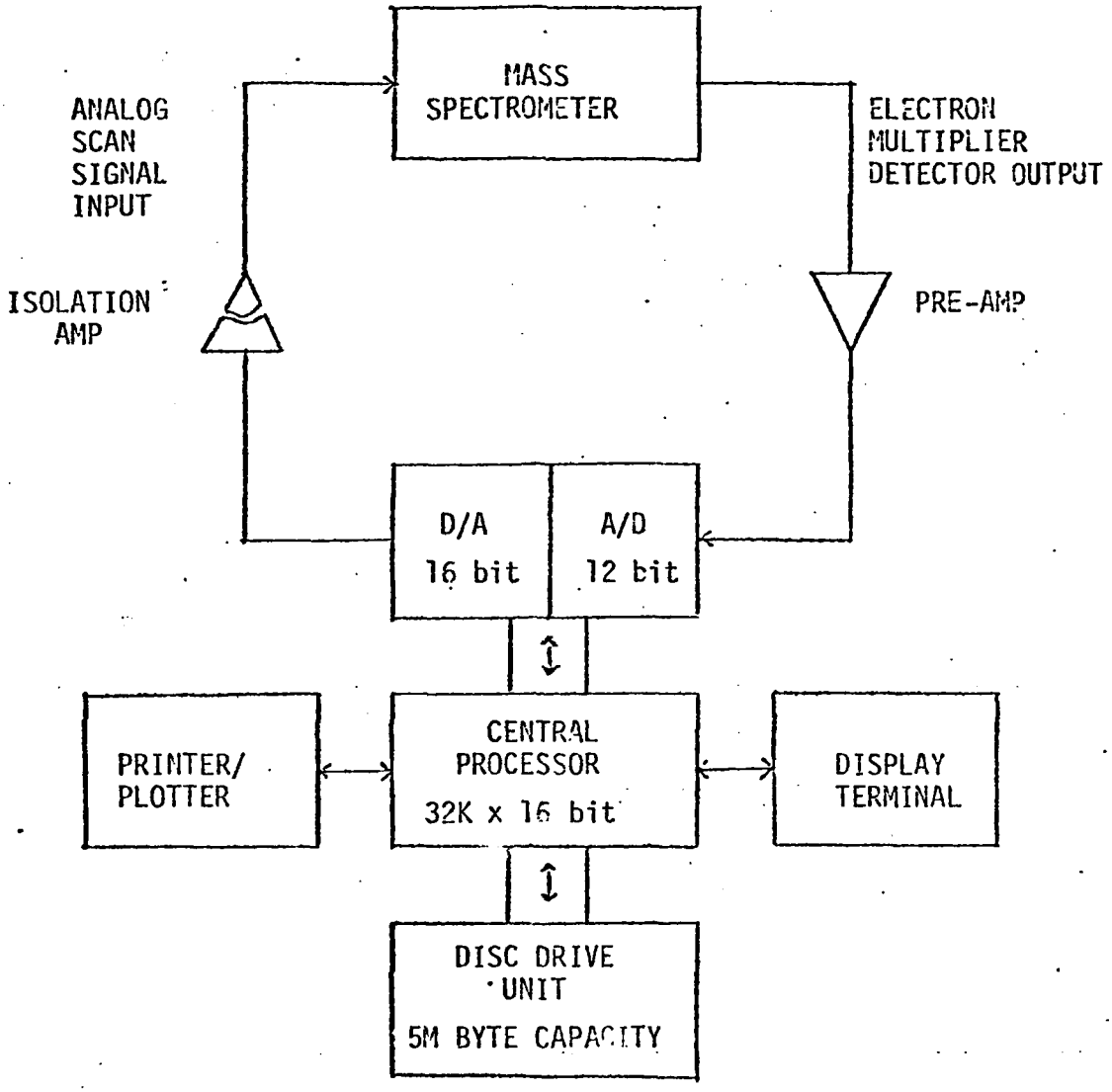
3. Computerized Data Acquisition

We have successfully interfaced each of our spectrometers to the FINNIGAN 2400 data system. Recently we have acquired and installed a dual acquisition system that allows simultaneous operation and data gathering on two mass spectrometers. The data system generates a digital scan function with selectable scan times and upper and lower mass limits. This digital scan function is applied to a 16-bit D/A converter to provide as analog signal to drive the spectrometer magnet. For use with the double focusing CID instrument this analog signal is applied to a current programmable magnet power supply (Alpha Scientific Model 3048) using an Analog Device model AD284J isolation amplifier. This configuration, comprising a computer generated digital scan, D/A converter and current programmable supply, gives highly reproducible scans on each instrument to which the computer is interfaced.

For field ionization operation, the accelerating voltage is monitored with a 4 1/2 digit digital voltmeter and maintained at the same value (nominal 5000V) at the beginning of each sample run to provide the same mass-to-time function for each analysis. The detector output is sampled by a 12 bit A/D converter with computer controlled integration times of 25 to 200 microseconds (see Fig. 3).

Mass assignment is achieved by a time-to-mass calibration curve established with a known mixture of reference materials. The calibration algorithm utilizes a higher order curve fitting, capable of accurate interpolation and extrapolation for mass assignment, provided the spectrometer

FIGURE 3



MASS SPECTROMETER-DATA SYSTEM BLOCK DIAGRAM

scan function is reproducible. In our system the digital scan function and the stable current programmed power supply gives mass assignment stability within 0.3 amu from sample to sample. Mass assignment drifts of less than 1 amu are generally maintained over several days.

In spite of the inherent mass assignment reproducibility we recalibrate the instrument for each sample. Calibration for field ionization is more difficult than with electron impact. Due to the absence of fragment ions, we must calibrate the mass range with a mixture of compounds with similar vapor pressures, in order to obtain spectra with all calibration peaks present in a single scan. Presently we use a mixture of seven compounds (see Figure 4) covering a molecular weight range for m/e 73 to m/e 298. Using an instrument with a nominal resolution of 500, the calibration with this mixture is generally accurate to 0.05 amu at mass 300. Assignment of masses outside the 73-298 amu range is by extrapolation of the computed calibration curve. Extrapolated assignment of masses up to 100 amu above m/e 298 is accurate and stable to $\pm$ 0.2 amu. This is measured by observing the assignment of a high M.W. peak (e.g. cholesterol MW 386) using several different calibration curves.

For analysis of urine extracts the mass spectrometer is scanned from 1 to 450 amu in 12 seconds. The total scan cycle time is 17 seconds including the time spent on returning to the starting mass and allowing for field stabilization. Samples are placed in the solid probe which is placed in contact with the ion source. Data acquisition starts as soon as the sample is in position near the ion source. The sample is maintained at 20^{o}C using air cooling of the probe during sample introduction, and following initial contact with the ion source which is maintained at 200^{o}C. The probe temperature is

FIGURE 4

MASS	INTENSITY	
73.0528	1000	DIMETHYL FORMAMIDE
109.0682	2600	ORTHO-PHENYLENEDIAMINE
152.0474	4000	PARA ANISIC ACID
170.0732	1400	PARA HYDROXYDIPHENYL
244.1253	800	TRI PHENYLMETHANE
250.1107	1000	METHAQUALONE
298.2872	400	METHYL STEARATE

MASS LIST DATA: 00201 # 4 BASE M.
10/08/79 12:38:09 1.00 CAL: 00201 # 4 RIC:
SAMPLE:

16 0.00 0.00 400. MINIMA MIN INTEN: 40
326 # 0 MAXIMA
MASS % RA % RIC INTEN.

MASS		% RA	% RIC	INTEN.
73.05	M	100.00	71.72	44992.
74.06		4.16	2.98	1870.
108.07		16.47	11.81	7409.
109.07	M	0.96	0.69	430
152.05		4.28	3.07	1924.
170.08		3.96	2.84	1782.
244.12	M	1.99	1.43	897.
250.11	FM	2.32	1.67	1046.
298.28	M	1.55	1.11	698.

TABLE 1: UPPER: LIST OF ACCURATE MASSES AND COMPOUNDS USED FOR CALIBRATION.

LOWER: LIST OF MASSES ASSIGNED BY CALIBRATION PROGRAM FOR SCAN TAKEN OF CALIBRATION MIXTURE.

linearly programmed for 20°C to 200°C increase over 20 min, and then held at 200°C for 5 minutes.

The data system records and stores 90 individual mass scans during the sample volatilization. These scans may be displayed in graphical or in tabular format during acquisition. The data is stored in an accurate mass format with the mass of each peak assigned to within $\pm$ 60 ppm amu by the current calibration file. At the end of the acquisition period, the individual scans are summed following conversion of the accurate masses to nominal masses, and this integrated spectrum is used in all subsequent data processing steps. Immediately following the urine sample, a new calibration sample is introduced and 10 scans acquired over a temperature increase from room temperature to approximately 70°C. Due to the high volatility of some components in the low molecular weight range, large changes in the relative peak intensities occur during the calibration run, and only the scans which have sufficient intensity of all components of the mixture are utilized for calibration.

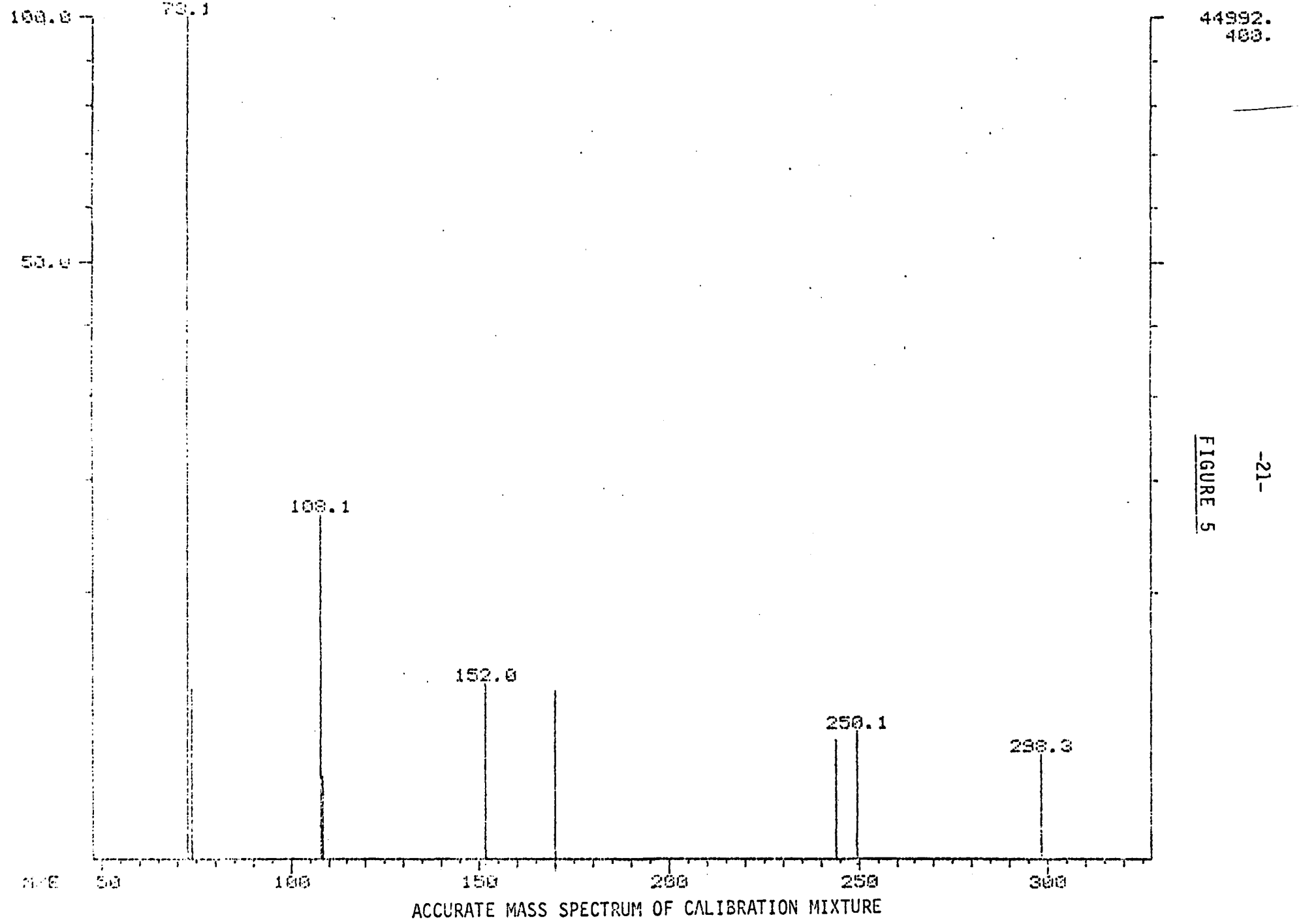
The drift in mass assignment of one of the components, methyl sterate, MW 298.28, is compared between this and the previous calibration file. This peak has the smallest intensity and the highest mass in the mixture, so that its mass assignment is subject to the largest variation due to ion statistics and scan irreproducibility. Differences in mass assignment for two consecutive calibration runs are typically within 0.1 amu.

After calibration the source and probe are baked out to remove any residue, in preparation for the next sample. The entire sequence of calibration, bakeout, and sample analysis requires about one hour. At least once each day the sensitivity of the ion source is measured by introducing 1 microgram of adenine, and acquiring data with the same scan sequence used for urine samples. The integrated signal for this sample is required to be greater than 5000 ions. Correcting for the percentage of scan time actually

spent monitoring the molecular ion, (10 ms/17 sec) this is equivalent to 10^{-12} Coulombs/microgram. Over the actual scan range employed this implies a minimum of 100 ions collected for 20 ng of a single component in an actual sample mixture. Examples of calibration and sample data are given in Figs. 5-9.

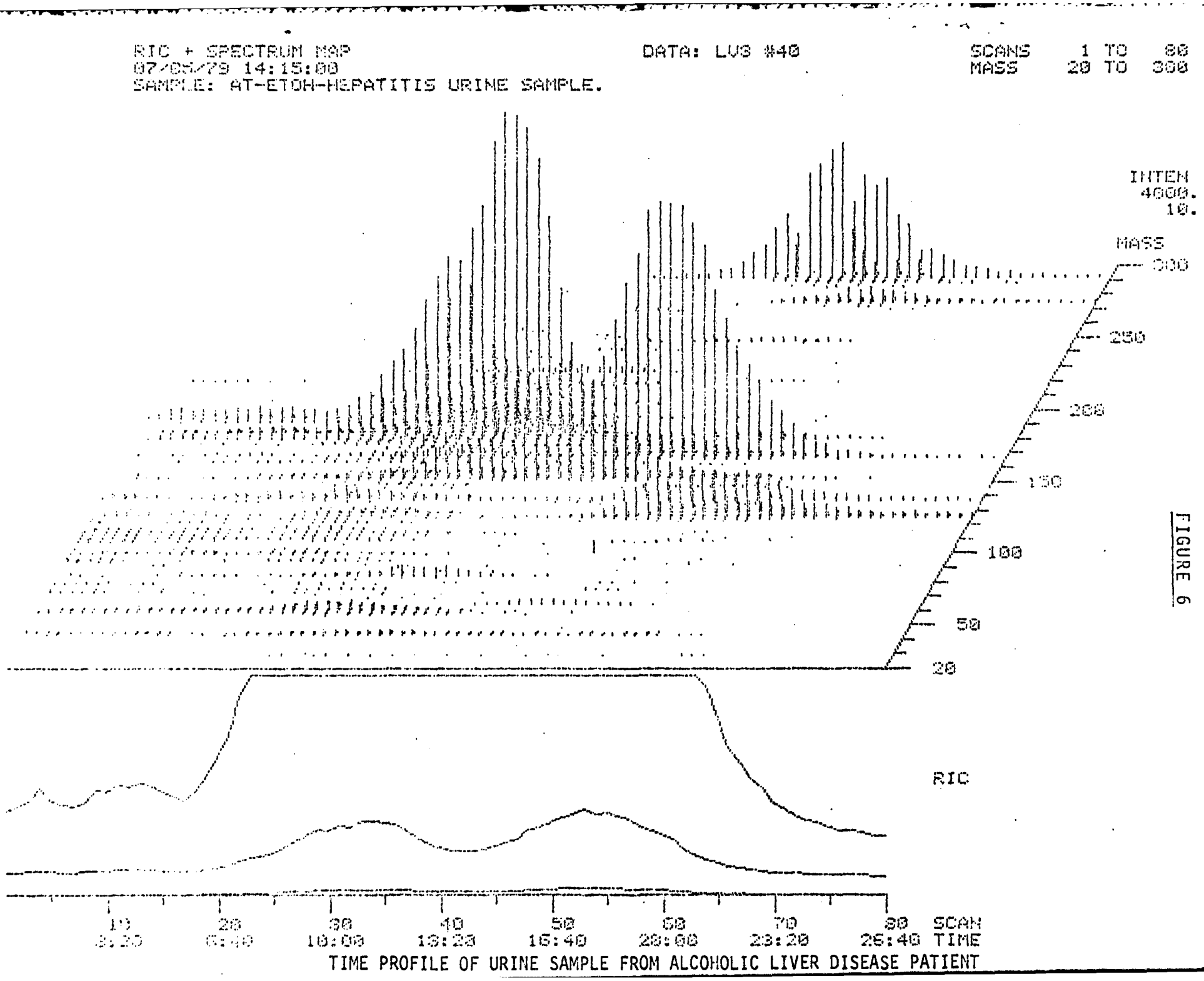
A standard urine extract is also analyzed at least once a week as an overall test of instrument performance. This sample is a 10 microliter aliquot of a of a concentrate obtained from the extraction of 20 ml of urine using a larger column appropriately scaled up in size to cope with a bulk sample.

Initially, profiles were recorded using a single focusing magnetic sector spectrometer (see 1979's report). In certain samples, particularly from patients with liver disorders, we encountered mass regions where ions appeared with a continuum of energies covering several amu of the mass scale. Further analysis revealed that many of these ions possessed energy-to-charge ratios greater than the main beam energy. These ions are a result of the formation of doubly charged species, which upon collision with neutral molecules either gain an electron or decompose to smaller fragments. In our instrument the most likely place for such collisions to occur is in the lens between the source and the object slit. This region has the highest density of neutral molecules as well as of focused deccelerated ions. The distribution of energies arises from the fact that these processes may occur anywhere between the ionizer and the object slit, in a region with a potential gradient. The final energy of the detached ion will be a function of the position where the transition from doubly charged to singly charged species occurred. The acceleration of the ion as a doubly charged species, even over a small fract' . of the net accelerating voltage, will result in an ion with a final



ACCURATE MASS SPECTRUM OF CALIBRATION MIXTURE

FIGURE 5



TIME PROFILE OF URINE SAMPLE FROM ALCOHOLIC LIVER DISEASE PATIENT

FIGURE 6

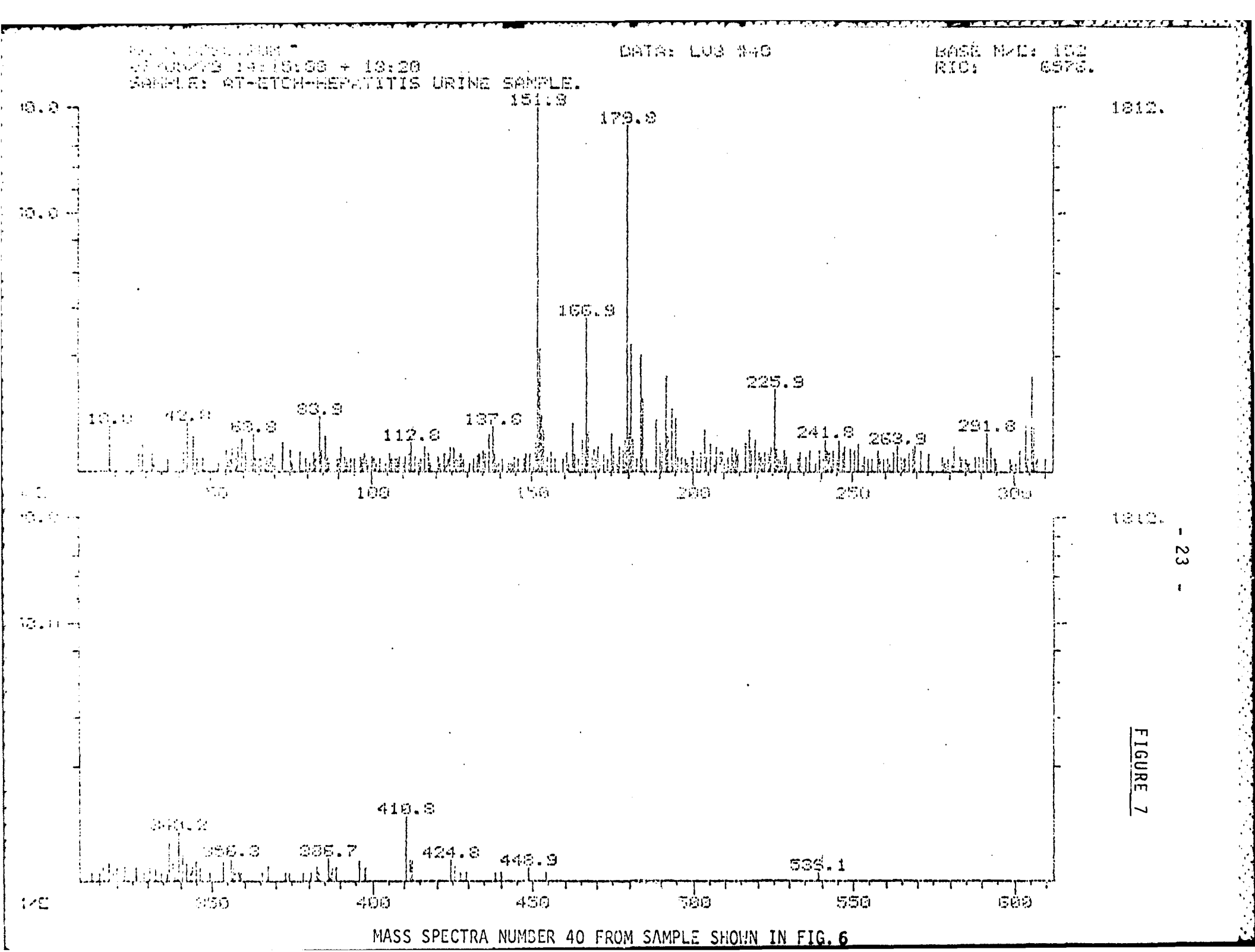


FIGURE 7

MASS SPECTRA NUMBER 40 FROM SAMPLE SHOWN IN FIG. 6

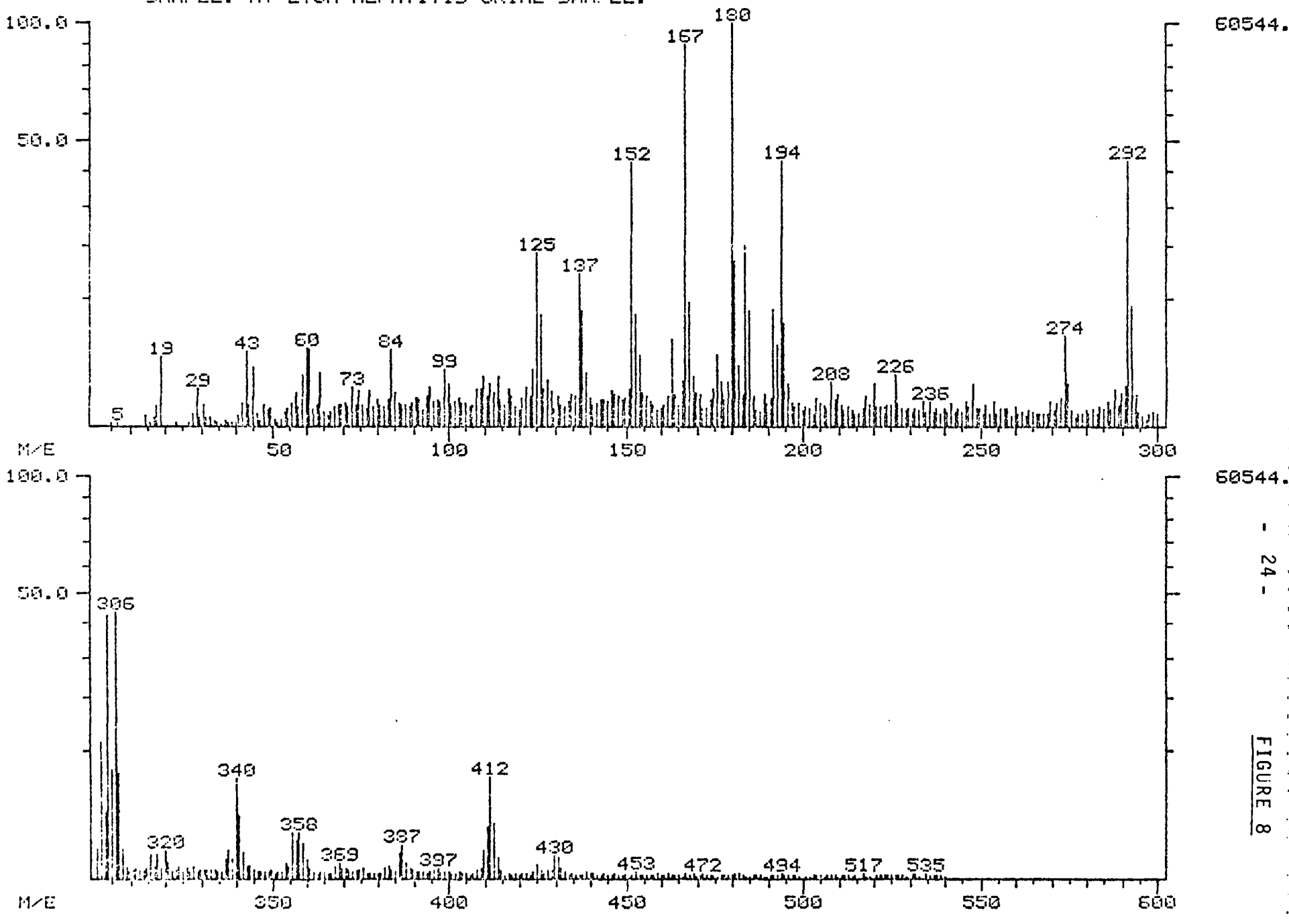

MASS SPECTRUM
07/05/79 14:15:00 + 0:20
SAMPLE: AT-ETOH-HEPATITIS URINE SAMPLE.
DATA: LVA3 #1
BASE M/E: 180
RIC: 438784.
100.0
50.0
60544.
M/E
50
100
150
200
250
300
180
167
152
194
292
125
137
274
19
43
60
84
99
29
73
208
226
236
5
100.0
50.0
60544.
M/E
350
400
450
500
550
600
306
340
412
358
320
387
369
397
430
453
472
494
517
535

FIGURE 8

FIGURE 9

MASS LIST DATA: LVA3 # 1 BASE M/E: 180
07/05/79 14:15:00 + 0:20 RIC: 438784.
SAMPLE: AT-ETOH-HEPATITIS URINE SAMPLE.

5 0.00 0.00 400. MINIMA MIN INTEN: 400.
540 # 500 MAXIMA

MASS	% RA	% RIC	INTEN.	MASS	% RA	% RIC	INTEN.
19.00?	2.97	0.41	1798.	181.00	16.86	2.33	10208.
29.00?	0.83	0.11	503.	182.00	2.26	0.31	1366.
43.00?	3.48	0.48	2108.	184.00	20.35	2.81	12320.
45.00?	2.05	0.28	1242.	185.00	8.26	1.14	5000.
57.00?	0.70	0.10	424.	192.00	8.56	1.18	5184.
59.00?	1.62	0.22	979.	193.00	4.02	0.56	2436.
60.00?	3.76	0.52	2276.	194.00	43.71	6.03	26464.
61.00?	3.52	0.49	2132.	195.00	6.43	0.89	3896.
64.00?	1.79	0.25	1084.	196.00	1.14	0.16	689.
73.00	0.92	0.13	556.	208.00	1.23	0.17	746.
75.00	0.80	0.11	485.	220.00	1.13	0.16	683.
78.00	0.80	0.11	486.	226.00	1.60	0.22	969.
84.00	3.63	0.50	2200.	248.00	1.10	0.15	669.
85.00	0.69	0.10	417.	274.00	4.96	0.68	3000.
95.00	0.91	0.13	552.	275.00	1.15	0.16	694.
99.00	1.94	0.27	1176.	288.00	0.89	0.12	538.
100.00	1.16	0.16	700.	290.00	0.68	0.09	410.
108.00	0.89	0.12	537.	291.00	0.99	0.14	602.
109.00	0.90	0.12	544.	292.00	43.45	5.99	26304.
110.00	1.53	0.21	926.	293.00	9.00	1.24	5448.
112.00	1.15	0.16	694.	302.00?	11.58	1.60	7008.
113.00	0.72	0.10	438.	303.00?	2.68	0.37	1624.
114.00	1.51	0.21	912.	304.00?	42.60	5.88	25792.
117.00	0.87	0.12	528.	305.00?	7.21	1.00	4368.
122.00	0.91	0.13	553.	306.00?	43.76	6.04	26496.
124.00	1.95	0.27	1180.	307.00?	6.83	0.94	4136.
125.00	18.50	2.55	11200	340.00?	5.98	0.83	3620.
126.00	7.78	1.07	4712.	341.00?	2.43	0.34	1470.
127.00	0.89	0.12	536.	356.00?	1.28	0.18	777.
128.00	1.36	0.19	821.	357.00?	0.95	0.13	577.
129.00	0.75	0.10	455.	358.00?	1.32	0.18	802.
137.00	14.24	1.97	8624.	359.00?	0.75	0.10	452.
138.00	8.09	1.12	4896.	387.00?	0.71	0.10	430.
139.00	1.74	0.24	1054.	411.00?	1.58	0.22	956.
146.00	0.81	0.11	493.	412.00?	6.40	0.88	3876.
151.00	0.90	0.12	542.	413.00?	1.84	0.25	1116.
152.00	42.86	5.91	25952.				
153.00	7.70	1.06	4664.				
154.00	3.13	0.43	1896.				
155.00	0.67	0.09	406.				
163.00	4.74	0.65	2872.				
166.00	1.31	0.18	795.				
167.00	89.32	12.32	54080.				
168.00	9.59	1.32	5808.				
169.00	1.55	0.21	940.				
170.00	0.68	0.09	414.				
175.00	0.88	0.12	534.				
176.00	3.03	0.42	1834.				
177.00	1.18	0.16	713.				
179.00	1.22	0.17	737.				
180.00	100.00	13.80	60544.				

PARTIAL MASS LIST OF INTEGRATED SPECTRUM SHOWN IN FIGURE 8

energy greater than the main beam energy. In order to obtain unit mass resolution of the singly charged normal molecular ions, a double focussing instrument employing an energy analyzer is required. Since the extent of these processes is variable and may occur in any sample, we now analyze all FI samples on the reverse geometry double focussing instrument (the CID instrument) described in the 1979 report.

Chemical ionization mass spectra were obtained on a DuPont 21-491B mass spectrometer equipped with a chemical ionization source using isobutane reagent bas at a source pressure of 0.3 to 0.5 torr. Mass resolution of 500 and a 10 sec cyclic scan from 1 to 550 amu gives integrated counts of 1.5 to 2 x 10^5 for 1 microgram of adenine evaporated from the probe, corresponding to 3 x 10^{-11} coulombs/microgram sensitivity. For analysis of urease digested urine samples the solid probe input power is programmed to evaporate the entire sample within 45 scans in 7.5 minutes. The Finnigan model 2400 data system is used also with this instrument to acquire and store spectra.

4. Data Reduction and Transmission

The mass spectra acquired are processed locally to facilitate transmission to the University's main computer (Control Data Corporation CYBER 173). A second set of operations in the CYBER then follows, and results in spectral data suitable for the subsequent statistical operations.

Within our dedicated Finnigan/INCOS computer system the individual mass spectrometer scans are stored as acquired. Upon completion of the run a single spectrum is produced by addition of all scans. This spectrum is then translated into a Fortran-readable format suitable for transmission to the CYBER main computer over a 1200 baud multiplexor phone link.

When the spectra are on file in the CYBER, format and simple logical checks are made on the data, which are then corrected as required. Each spectrum is then normalized to unit total area, with individual peak areas greater than 5 per cent of the total excluded from the normalization sum (for the rationale see 1979 report). All statistical analysis is then accomplished using these normalized spectra.

C. DIAGNOSTIC STATISTICAL ANALYSIS.

1. Introduction

During this second phase we have devoted a substantial effort to evaluate our diagnostic statistical analysis comparing a number of alternative techniques for the selection of variables and for the separation of cases into diagnostically meaningful groups.

The statistical analysis in this project has the following objectives:

a. To separate the biological samples (cases) into statistically distinct groups correlated with the disorder of interest (e.g. healthy vs. diseases, healthy vs. alcoholic liver disease, pneumonia vs. bronchitis, bacterial vs. viral pneumonia, etc.) by a characteristic set of variables.

b. To assign correctly an unknown case to one of a number of pre-specified groups, which have been previously developed on the basis of a continuously increasing learning set.

c. To identify the variables that best characterize a given pathological state, in order to facilitate the understanding of the biochemical nature of the disease and possibly also to explore the possibility of quantitative assay of the particular metabolite by a simpler non-mass spectrometric analytical technique.

d. To identify variables that co-vary, for two reasons - first to attain a better understanding of the biochemical nature of the disease, and second to minimize undue weighting bias in the differentiation of cases into diagnostic groups.

Different multivariate analysis techniques have provided us with answers to these questions with different degrees of efficiency. As stated above, we have started to compare different statistical techniques for their merits in meeting our objectives. In the following sections we shall discuss the current status of this evaluation, which will be continued during the forthcoming stages of this project.

We may separate our current statistical analysis into two phases - the selection and rating of variables (mass spectral peaks) according to their diagnostic value, and the classification of cases (patient' urine or tissue culture media) into groups.

2. Selection and Rating of Variables

Each normalized mass spectrum comprises hundreds of variables (peaks) each of which represents the concentration of a metabolite (or a group of metabolites sharing the same nominal mass) in the biological sample. When we compare the magnitudes of a given variable in samples coming from two biochemically distinct groups we observe three types of variation:

- variation due to the analytical procedure
- variation due to biological variance (due to genetic or nutritional factors), and
- variation associated with the experimental difference between the groups, e.g. variation due to the pathological status of a human subject, or the infected status of a tissue culture.

Ideally there should be no variation of the first kind and the variation due to the pathological status should be by far larger than the biological variation. In reality, however, the majority of variables show a large biological variation (on top of a finite experimental variance) and are thus of minimal diagnostic value (we shall call these diagnostically useless variables). Our problem is, therefore, to select those variables which may be diagnostically useful.

Any statistical diagnostic procedure will become ineffective if given an excessive number of diagnostically useless variables, even in the presence of many useful ones. On the other hand, we would like to utilize every useful variable since each of these increases the diagnostic power of the classification procedure. An acceptable variable selecting procedure has, therefore, to identify and reject useless variables, while retaining all the useful ones. Moreover since there will always be variables more "useful" than others, an adequate procedure should rate them accordingly, and thus allow us to optimize the diagnostic power of the diagnostic procedure.

There is another factor that should be taken into account - covariance. In a biological system there are many variables that are biochemically interrelated, so that their variation associated with a given pathological state is interdependent. If such variables are used in a statistical method based on a pattern of a given number of independent variables, they may bias the result by giving an undue high weight to a single variation (accompanied by a set of dependent variables). It is important, therefore, to identify such co-variances and eliminate the satellite variables from the diagnostic classification pattern. A desirable feature of a selecting technique would be, therefore, the ability to identify covariance and minimize its effect.

a. The Wilcoxon Test

This non-parametric ranking of variables according to the probability of being constituents of the same population has been described in our previous reports. This ranking has two important shortcomings: first, it does not identify covariance, and second, it will not discriminate against artifactual deviants. In fact this treatment may give a variable (peak) with an unlikely large deviation in one of the cases (spectra) an undue outstandingly low probability. This second shortcoming is avoided in the t-test variable rating.

b. The t-test Rating

We have applied the t-test program P3D of the BMDP programs package (UCLA 1977) to determine the null hypothesis that each of the variables in two tested groups of cases belongs to the same population. Unlike in the Wilcoxon test this probability is calculated on the basis of deviations from the group's average, while taking into account the difference in values between the two group averages. This program also provides us with the variance of each variable in each group, so that artifactual deviants can be readily identified and discounted.

In spite of the intrinsic differences between the Wilcoxon and the t-test, the two programs identified 50 out of 400 peaks from the same test set of spectra as of prime diagnostic value with an overlap of over 90% of the variables selected, (although the order of ranking by the two procedure was somewhat different). In view of this finding and since the t-test provides additional useful information we prefer now to use this program for selection of the diagnostic peaks.

c. The Stepwise Discriminant Analysis Procedure.

This procedure to be discussed below, selects and rank orders variables according to their "F" values. The F value for each variable is proportional to the square of the intergroup difference and inversely proportional to the square of the intragroup variance around that group's average. Since this procedure requires considerably more computer capacity than the two preceding methods, it can handle just a limited number of variables (50 peaks in our case). This limitation requires pre-selection of variables by one of the preceding techniques, to be followed by their F value ranking according to their usefulness in separating the cases into distinct groups. The main feature of this statistical procedure is its ability to identify and reject co-variants. In spite of this important advantage, the discriminant analysis can hardly be considered a practical peak selecting procedure, because of its high demand on computer time and capacity. However since this technique is being used as a group classification and case assignment procedure, the peak selection and ranking according to the F values may be considered as a fringe benefit.

d. Modes of Use of Selected Variables

The variables for a given classification procedure can be selected by virtue of meeting a certain arbitrary criterion (e.g. having a p value below a given value), or by rank ordering according to a given criterion (e.g. starting with the variable of lowest p value, followed by the next lowest, and so on) and then picking an arbitrary number (e.g. 50) having the lowest p values. The selected variables can then be used in the group classification procedures without any weighting.

Alternatively a characteristic classification parameter (e.g. the p value) can be used as a weighting factor. Since the diagnostic usefulness increases as p decreases, using 1/p as a weighting factor is perhaps the simplest weighting procedure. This weighting will, however, give variables with very small p values a very high weight. Alternative weighting factors could be for instance $1/p$, $1/\sqrt[3]{p}$, or $\log 1/p$, which would decrease the overweighting of variables with very small p values.

There are advantages to either peak selection procedure. The discriminant method (by an arbitrary cut-off) requires human judgement for each set of cases. This shortcoming can be eliminated by rank order cut-off. Using the t-test one can use the variance in addition to the t or p values as a second criterion in selection.

The weighting procedure while free from subjective intervention nevertheless requires optimization to obtain the best use of separating variables. However, the weighting of variables as in the WNI procedure (see our 1979 report) involves the choice of an arbitrary weighting function which is at best a compromise between an optimized use of the variables and a use based on a subjective threshold.

3. Group Classification Procedures

During the previous phase of this program we have used basically just one classification procedure, namely the weighted non-parametric index (WNI) method which has been described in our previous reports.

This procedure which is simple and straight forward has certain limitations. First, it is applicable to only two groups of variables. Although when WNI (1) is plotted vs. WNI (2) one can obtain some secondary clustering, indicating sub-groupings, but the separation between these is biased by the original choice of two groups, since the average value of each variable in each group is the reference point for the WNI calculation. Second when small sets of cases are analysed, the WNI values are strongly influenced by variables with large variances from their respective group averages. This is especially true when the difference between the group average is relatively small. Third, the diagnostic referent point on the D = WNI (1) - WNI (2) scale is arbitrary, which becomes problematic if the D values for members of the two groups form a progressive continuum without a significant gap between D's of the two groups. Fourth, this "non-parametric" procedure does not provide us with a measure of probability of a given case belonging to each group, thus it is lacking a quantitative measure for the diagnostic assignment of a given case to a particular group.

In view of these limitations, we have experimented with two other classification procedures - the clustering analysis procedure (P2M procedure BMDP UCLA, 1977) and the stepwise discriminant analysis procedure (P7M BMDP UCLA, 1977). The former classification is free of the bias of assignment of cases to a particular number of groups, whereas the latter can handle efficiently a large number of pre-specificed groups and it provides the probability of each case belonging to any of the groups in question.

a. The Clustering Analysis

In this procedure one represents each of the cases (each with n variables) as a point on a surface of a n dimensional space. This is done by calculating an n dimensional vector as a resultant of the values of the variables measured on n orthogonal coordinates. If we have many cases, each constituting a point on the n dimensional surface, we can calculate the distance on this surface between any given two points. The clustering procedure selects the two points with the shortest n dimensional Euclidian distance between them, producing a cluster of two. Then the program tries to find among all the remaining points a (third) point (case) closest to the first two points, forming a cluster of 3. Again the program selects and registers a (fourth) point closest to the cluster of three, and so on.

When the distance between the growing cluster and the next point is larger than between a pair of the remaining points, a new cluster of two is selected, which can again grow, aggregating points in its vicinity. This process continues until the nearest distance remaining is the distance between the boundarie of two clusters, which are then registered as a cluster of clusters. The classification is ended when all points are accounted for, when a master cluster containing all the points (all the cases) is registered.

This approach is completely bias-free as far as the number of groups (clusters) it will form from a set of cases; only after clustering can one check the a priori assignment of a given case against the cluster it ended up in. The program also allows the identification of the relative positions of points (cases) within clusters. On the other hand, the procedure does not test the variables for their variances from a group average (which is done in the WNI and in the discriminant analysis classifications) or for co-variance, which is performed in the discriminant analysis.

b. The Stepwise Discriminant Analysis

This classification procedure separates the cases into a prespecified number of groups after analyzing the variance of each variable. This procedure also selects those variables which separate the cases into the specified groups most effectively.

The procedure first determines the variance of each variable within each group and compares it to the variances between groups. The comparison is done by calculating F values, i.e. dividing the square of intergroup variance (S_G) by the square of the variances (S_v) of the individual variables around the corresponding group average: $F = S_G^2/S_v^2$. The program then selects the variable with the highest F value and if this is larger than a pre-specified threshold "F-to enter" it will use this variable to classify the cases into groups.

In the second step it will test all remaining variables for their ability to separate between groups on a 2 dimensional surface, comparing again the new intergroup variance to the group variances. The variable with the highest "F to enter" value at this step will be then added to the first selected variable, provided its F value exceeds the "F to enter" threshold and provided that a correlation coefficient between the two variables is not above a specified limit. A high correlation would obviously invalidate the group classification by two presumably independent variables.

After the two classifying variables are selected, the program computes for each case a point (vector) on a 3 dimensional surface using each of the remaining variables combined with the 2 variables selected in steps 1 and 2. The variable with the highest F value is selected, provided it exceeds the threshold and that it does not co-vary with the combination of the two first variables. If it does not fulfill the second condition the variable with tr

next highest F value is selected and tested against the two criteria. The variable selection procedure is continued until all remaining variables end up with F values below the threshold or exhibit excessive covariance with the set of selected variables.

Following each step the program also re-checks the F values of all the variables selected for classification up to that step, since their F values will change with each added variable. If any of the previously selected variables has an F value lower than a given threshold ("F to remove") it will be removed from the classification set, and the procedure is repeated again for all the remaining non-classifying variables to select a new one with an acceptable F value and covariance coefficient. The variables selected at each step are combined to form a linear, optimized classification function (n-dimensional vector) that maximizes the S_G^2/S_V^2 ratio.

Once the variable selection process is complete using say m variables, the cases are classsified into groups as points on a m dimensional surface and the intercase distance is calculated. From these the centroids or group averages are computed as is the probability of assignment of each case to any given group. The grouping can be presented graphically in two dimensions as points indicating the groups centroids, or as clusters of points (representing the individual cases) around the centroids.

4. A Comparative Evaluation of the Statistical Classification Procedures.

As stated above each of the three classification procedures tested by us has more desirable and less desirable features when compared to the others. To illustrate this we applied the three procedures to the same set of data, namely analysis 39 spectra from urine of 14 patients with alcoholic liver disease compared with 26 spectra from 13 healthy adults.

The results of the Wilcoxon test on 100 peaks of lowest p values is given in Figure 10. The WNI values of the 65 cases for the two groups using 1/p weighting as well as the values of D = WNI (1) - WNI (2) are given in the same figure. One can see here that all the D values of the pathological samples (LV) are negative and smaller than -35, whereas those of all the controls (CP) are positive, with the exception of case CP22 which is negative (-9.5) but still significantly larger than any of the LV's. In other words this classification did not show any overlap between the tested groups. A computerized graphical presentation of the WNI data is presented in Figure 11 where each case is presented by its WNI (1) and WNI (2) values. We see here the clustering of the two groups with a distinct region of demarcation between them. Although WNI (1) = WNI (2) for a case would indicate that it equally belongs to the two groups, this is not necessarily true when applied statistically to two groups where at least one cluster has D values very different from zero.

The clustering analysis applied to the same cases is presented in Figure 12. Here cases 1 to 23 were of liver patients, and cases 24-48 were of controls. We see here that clustering began with case 17, 13 and 6 followed by cases 19 and 11. By the end of the clustering process all samples 1 to 24 (the pathological cases) were in one cluster with case 44 (of the controls) being the next one added to this cluster (but only in step α41 in the amalgamation order). All the other controls are again in a distinct different cluster. One may also distinguish some sub-clusters like cases 1, 2, 10 and 11 or 15, 21, 17, 16, and 20 among the pathological samples. Since this program does not presume any predetermined groups for classification these subgroups may have some biochemical features in common in addition to being part of a liver disease population.

FIGURE 10

161	.000000001	116	.000000055	231	.00006948	192	.00065667	245	.00326195	170	.00884791
156	.000000001	146	.000000276	81	.00007691	18	.00065667	185	.00348539	219	.00953402
197	.000000002	168	.000000376	136	.00008504	200	.00071093	122	.00372209	241	.01051855
339	.000000005	160	.000001104	118	.00013814	202	.00083160	147	.00397270	201	.01051855
85	.000000007	219	.000001135	145	.00015174	148	.00089852	205	.00397270	149	.01051855
79	.000000012	112	.000001135	188	.00015174	152	.00089852	285	.00451870	247	.01051855
150	.000000016	234	.000001614	175	.00016997	265	.00112901	153	.00461546	216	.01051855
121	.000000018	194	.000001713	142	.00021884	141	.00112901	77	.00512908	226	.01113220
140	.000000023	86	.000001949	211	.00023920	31	.00151755	285	.00546034	181	.01245139
224	.000000028	169	.000001988	318	.00026140	163	.00151755	230	.00546034	227	.01245139
238	.000000030	233	.000001416	138	.00026140	165	.00163155	54	.00546034	71	.01393097
337	.000000032	340	.000001668	139	.00031121	105	.00163155	120	.00581003	287	.01467770
117	.000000034	78	.000002057	218	.00040184	240	.00175337	61	.00617901	229	.01549071
96	.000000042	207	.000002057	319	.00043691	29	.00188255	350	.00617901	215	.01634130
94	.000000044	210	.000002319	137	.00043691	178	.00232323	190	.00697827	179	.01634130
133	.000000050	43	.000002608	159	.00051535	73	.00240013	97	.00741037	238	.01634130
164	.000000059	139	.000003669	228	.00051535	246	.00285258	251	.00786536	217	.01634130
195	.000000061	60	.000004566	250	.00055910	187	.00305123	88	.00834420	101	.01634130
182	.000000064	183	.000005082	74	.00060614	193	.00305123	233	.00884791	303	.01723077
169	.000000065	205	.000005082	225	.00060614	190	.00326195	213	.00884791	242	.01816045

TAG	WNI(1)	WNI(2)	D
LV1	22.2279	61.8611	-39.6332
LV2	19.8587	59.4314	-39.5627
LV3	16.7556	70.1935	-53.4249
LV4	31.2241	77.4078	-46.1838
LV5	29.8909	79.4271	-49.5361
LV6	36.5374	80.3291	-43.7917
LV7	29.9554	79.3170	-49.3615
LV8	21.3087	59.6871	-38.3783
LV9	17.7453	64.2972	-46.5518
LV10(X5)	26.0268	62.7830	-36.7561
LV11(X6)	21.7947	63.7182	-41.9235
LV12(X1)	12.6833	69.4026	-56.7193
LV13(X2)	16.4550	71.2143	-54.7593
LV24	8.8966	70.2593	-61.3607
LV21	18.4758	63.9014	-45.4256
LV22	10.0309	68.4761	-58.4362
LV23	11.3226	72.3791	-61.0556
LV24	20.2200	78.8498	-58.6298
LV25	23.4524	76.9839	-53.5315
LV26	19.1874	59.9024	-40.7150
LV27	13.8303	65.2102	-51.3719
LV28	15.4991	65.7579	-50.2588
LV29	11.7760	67.0595	-55.2835
CP0	52.9444	28.4967	24.4478
CP1	61.3937	17.4223	43.9715
CP2	57.0847	37.8311	19.2536
CP3	51.9014	33.0917	18.8097

LV25	23.4524	76.9839	-53.5315
LV26	19.1874	59.9024	-40.7150
LV27	13.8303	65.2102	-51.3719
LV28	15.4991	65.7579	-50.2588
LV29	11.7760	67.0595	-55.2835
CP0	52.9444	28.4967	24.4478
CP1	61.3937	17.4223	43.9715
CP2	57.0847	37.8311	19.2536
CP3	51.9014	33.0917	18.8097
CP4	54.5366	29.2925	25.2441
CP5	54.6431	25.0456	29.5976
CP6	54.0397	30.2827	23.7570
CP7	68.1072	12.8667	55.2406
CP8	49.8786	37.7411	12.1375
CP9	61.3780	40.4931	20.8849
CP10	79.9826	21.5971	49.3855
CP11	69.6015	10.8920	58.7094
CP12	51.2427	33.2005	18.0422
CP13	52.3912	30.4874	21.9038
CP14	56.7367	21.7985	34.9383
CP15	55.6018	23.1143	32.4875
CP16	51.1059	29.3202	21.7858
CP19	47.4977	35.9974	11.5003
CP20	43.4696	33.4420	10.0277
CP21	55.9613	28.9414	27.0199
CP22	37.5939	47.0697	-9.4758
CP23	54.6401	10.1939	45.4462
CP24	53.3530	34.9375	28.4155
CP25	72.6593	46.5704	26.0848
CP26	73.5459	24.6155	48.9304

FIGURE 10

FIGURE 11

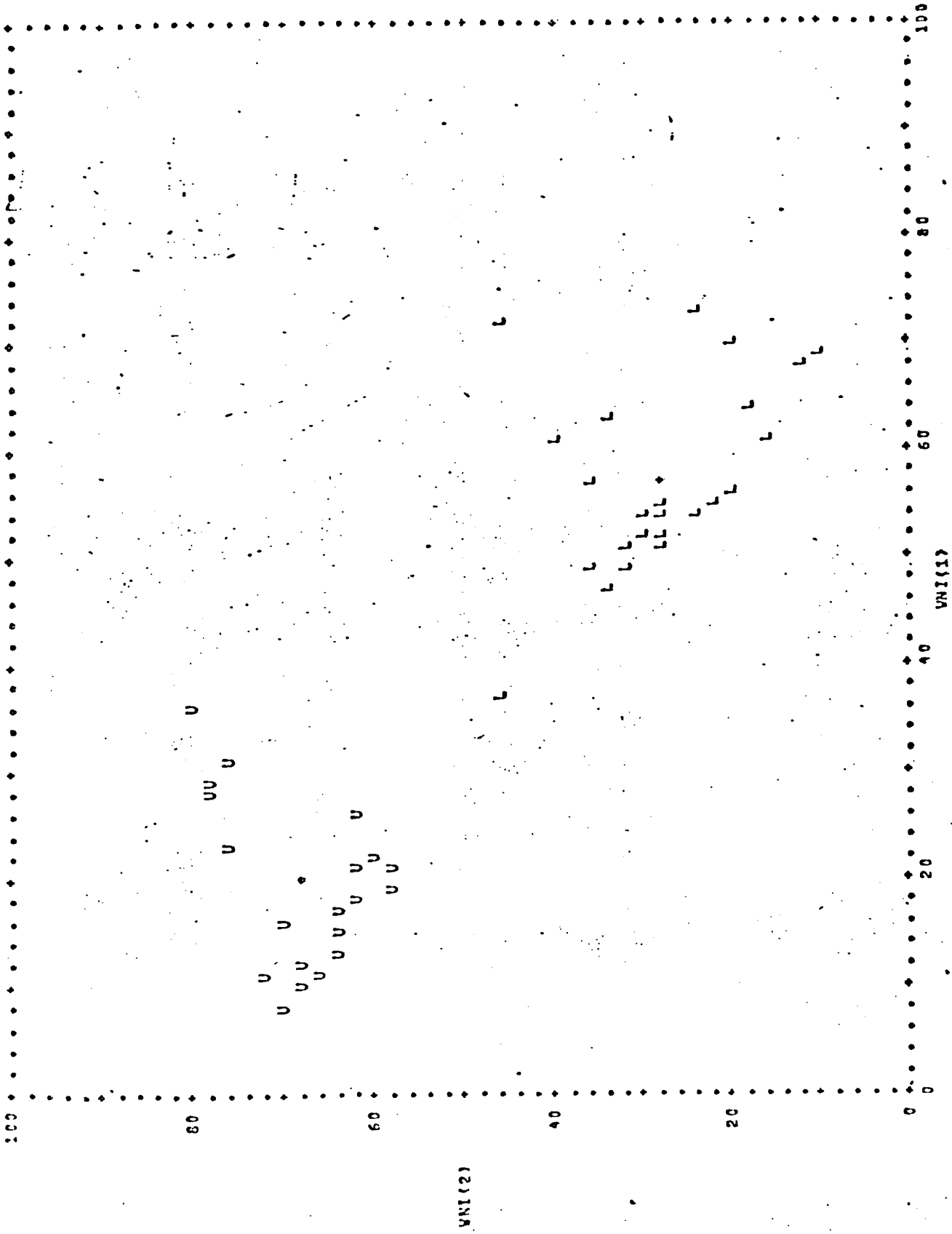


FIGURE 12

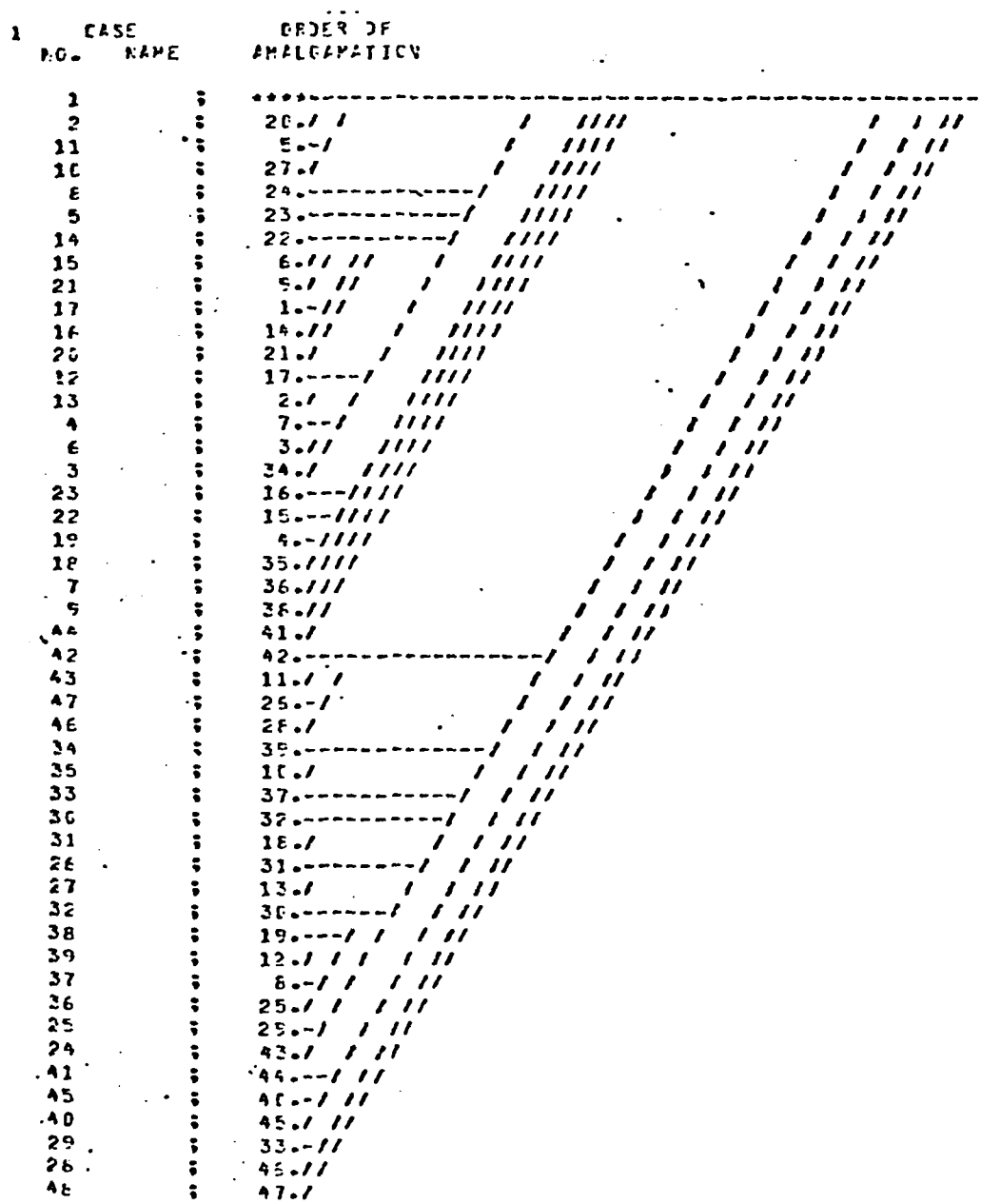


Figure 13 presents part of the computer output of the Euclidian distances between the points representing each case on an n dimensional surface. (The distance between 48 points to just 15 points of other cases are shown in this figure). These distances are then amalgamated or clustered as discussed above.

Figure 14 presents the same variables and their respective intragroup averages used for the subsequent stepwise discriminant analysis. Figure 15 shows the first two steps of selection of the two first variables for classification, namely mass peaks 197 and 95 respectively. The same figure then describes step #15 of the procedure by which 15 peaks were selected by the "F-to-enter" and covariance criteria. Figure 16 presents a later stage in the procedure - step #40, when just 30 variable (peaks) were selected as parameters in the classification, indicating that 10 variables originally selected were subsequently rejected due to F values below the ("F-to-remove") threshold.

Figure 17 presents the distances from each group centroid to the points representing each of the cases calculated from the 30 variables finally selected. This table gives also the probability of assignment of each case to a given group. One can see that each case in the L group have been assigned to it with a probability of unity and similarly each case in the C group (controls) has been assigned to this group with a probability of unity.

Figure 18 is a graphical presentation of the same data recalculated for 2 dimensional projection. The digits 1 and 2 are the positions of the centroids of each group respectively. The resolution of the computerized printout is limited, so that if more than one case fall on the same overall unit area on the plot they will be presented by just a single mark. Therefore, only 19 points are shown for the pathological sample and coincidentally 19 points were printed for the controls. The actual coordinates for each of the 48 cases is presented in Figure 17.

FIGURE 13

INITIAL DISTANCES BETWEEN CASES

CASE NUMBER	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.00	4.92	8.51	9.51	7.28	10.43	9.49	8.82	9.58	5.76	6.22	6.90	7.62	8.84	9.71
2	4.92	0.00	8.11	9.21	6.78	10.04	8.71	7.44	9.58	5.81	6.25	6.81	7.48	7.99	8.59
3	8.61	8.11	0.00	2.78	6.26	4.18	6.95	5.46	7.06	7.72	5.89	4.30	4.41	5.46	6.83
4	9.51	9.21	2.78	0.00	7.18	2.35	7.08	6.86	7.61	8.26	6.50	4.76	4.66	5.97	7.64
5	7.28	6.78	6.26	7.18	0.00	7.84	5.36	6.05	8.17	6.35	5.90	5.27	5.45	5.94	6.34
6	10.43	10.04	4.18	2.35	7.84	0.00	7.67	8.14	8.22	8.88	7.16	5.69	5.58	6.48	8.00
7	9.49	8.71	6.95	7.08	5.36	7.67	0.00	8.24	9.09	8.17	7.53	6.58	6.69	6.93	8.00
8	8.82	7.44	5.46	6.86	6.05	8.14	8.24	0.00	6.65	8.56	7.72	6.27	5.83	5.88	5.93
9	9.58	9.58	7.06	7.61	8.17	8.22	9.09	6.65	0.00	9.38	8.36	7.36	6.47	5.49	5.93
10	5.76	5.81	7.72	8.26	6.35	8.88	8.17	8.56	9.38	0.00	3.13	5.70	6.59	7.98	9.16
11	6.22	6.25	5.89	6.50	5.90	7.16	7.53	7.72	8.36	3.13	0.00	4.25	5.10	6.82	8.33
12	6.90	6.81	4.30	4.76	5.27	5.69	6.58	6.27	7.36	5.70	4.25	0.00	2.01	4.68	7.02
13	7.62	7.48	4.41	4.66	5.45	5.58	6.69	5.83	6.47	6.59	5.10	2.01	0.00	4.35	6.49
14	8.84	7.99	5.45	5.97	5.94	6.48	6.93	5.88	5.49	7.98	6.82	4.88	4.35	0.00	3.29
15	9.71	8.59	6.83	7.64	6.34	8.00	8.00	5.93	5.93	9.16	8.33	7.02	6.49	3.29	0.00
16	8.75	7.51	6.72	7.37	6.24	7.76	7.69	6.95	6.89	8.77	7.73	6.03	5.71	4.72	4.84
17	8.50	8.01	6.19	6.89	6.03	7.68	7.44	7.13	6.83	8.49	7.40	5.70	5.31	4.66	5.02
18	9.42	7.93	7.72	8.13	6.64	8.54	6.95	8.32	9.73	7.27	7.37	7.44	7.07	7.32	7.89
19	9.24	7.85	7.36	8.01	6.27	8.63	6.28	7.78	9.31	7.65	7.39	7.15	7.60	6.65	7.21
20	7.69	6.65	6.37	7.39	6.44	8.01	7.88	6.80	8.76	8.01	7.10	6.55	7.01	5.99	6.11
21	8.39	7.05	6.25	7.02	6.24	7.52	7.27	6.32	7.29	8.36	7.45	6.24	6.02	3.97	3.68
22	10.82	9.10	8.51	9.38	8.45	9.52	8.49	9.29	10.99	9.29	8.97	8.77	9.10	7.74	8.11
23	8.99	7.43	8.01	8.90	7.04	9.53	7.99	8.26	10.06	8.20	7.98	7.58	8.09	6.75	7.02
24	10.09	8.54	8.79	10.32	8.76	11.38	10.75	7.12	10.72	10.81	10.28	9.43	9.56	9.42	9.29
25	11.06	10.35	8.83	10.32	9.50	11.18	12.05	7.97	11.59	12.05	11.22	10.10	10.40	9.94	9.39
26	12.26	10.73	10.47	11.91	11.09	12.77	12.89	9.12	11.47	12.78	11.90	10.83	10.50	9.64	9.43
27	10.76	9.21	8.26	9.77	8.88	10.68	10.53	7.04	9.53	11.06	10.09	8.68	8.64	8.01	8.28
28	14.76	13.24	12.56	14.42	13.32	15.29	14.89	10.84	14.43	15.05	14.38	13.54	14.04	13.91	13.42
29	13.09	12.10	11.59	13.30	12.07	14.08	13.85	10.13	13.54	13.63	13.07	12.58	12.86	12.76	12.47
30	12.66	11.17	10.21	11.92	11.55	13.21	13.23	9.77	11.86	13.51	12.64	11.39	11.47	11.03	10.64
31	12.24	10.73	9.72	11.11	10.40	12.26	12.17	8.23	10.77	12.37	11.42	10.15	10.20	9.22	9.19
32	11.49	9.35	9.27	10.82	10.24	11.79	11.90	7.65	10.64	11.68	10.58	10.12	10.19	9.85	9.92
33	13.52	11.40	11.51	12.82	12.65	13.92	14.07	9.85	12.74	13.70	12.78	12.30	12.34	11.21	11.11
34	13.01	10.83	10.53	11.75	11.98	12.89	12.71	9.18	11.59	12.88	11.93	10.81	10.94	9.91	10.02
35	13.72	11.13	10.80	11.80	12.04	13.00	12.90	9.76	12.04	13.30	12.31	11.32	11.38	10.60	10.97
36	11.30	9.38	9.17	10.96	10.03	11.59	11.76	7.46	11.00	11.57	10.79	10.04	10.28	10.27	10.19
37	10.96	9.35	8.74	10.22	9.57	11.48	11.41	7.13	10.57	11.11	10.23	9.31	9.57	9.42	9.50
38	11.39	8.76	10.31	11.87	10.41	13.06	12.33	7.98	11.61	11.64	11.07	10.73	10.94	10.52	10.27
39	11.41	9.15	9.42	11.00	10.14	12.25	11.51	7.41	11.01	11.56	10.91	10.06	10.38	9.82	9.59
40	12.30	10.61	10.51	12.08	11.67	13.26	13.07	9.13	12.53	12.84	12.42	11.27	11.46	11.98	11.80
41	13.91	11.63	12.37	14.32	13.01	15.11	14.92	10.89	14.39	14.23	13.96	13.20	13.55	13.58	13.39
42	14.75	12.54	11.92	13.01	11.89	13.75	13.31	10.17	12.18	14.06	13.19	12.13	11.93	10.02	9.71
43	12.79	10.63	9.74	10.84	10.23	11.62	11.50	8.42	10.16	12.19	11.15	9.88	9.67	7.97	8.15
44	9.74	8.41	6.53	7.50	8.23	8.70	9.23	6.96	9.66	9.35	8.50	6.78	7.02	6.05	9.29
45	12.77	11.50	11.18	12.59	12.18	13.62	13.71	10.30	13.34	13.52	12.94	11.54	12.12	12.64	12.57
46	13.18	11.21	10.54	11.77	10.24	12.38	11.87	9.07	11.33	12.63	11.68	10.44	10.51	9.21	8.99
47	14.16	13.65	14.04	15.20	13.62	15.76	15.11	12.43	14.31	15.50	14.59	13.77	13.91	12.39	11.90
48	15.22	13.14	13.42	14.30	14.54	15.22	15.49	12.25	14.16	15.24	14.23	13.57	13.60	12.38	12.64

FIGURE 14

MEANS		
GROUP =	L	C
VARIABLE		
1 M74	.79695	1.61160
2 M78	1.41739	3.29560
3 M79	.96739	1.89840
4 M80	.95345	1.82760
5 M85	1.03348	1.79080
6 M93	.53935	1.67200
7 M94	5.37348	31.50440
8 M95	2.57522	12.10400
9 M96	1.14343	2.22920
10 M102	1.37261	.70400
11 M103	2.45304	1.13520
12 M109	1.15957	3.53600
13 M112	6.35000	2.52320
14 M116	1.57435	.75960
15 M117	2.53435	1.25240
16 M118	1.16609	.79560
17 M121	.73130	1.51080
18 M135	1.73595	2.63080
19 M140	1.89391	1.92440
20 M142	1.02304	1.54560
21 M145	2.73217	5.39840
22 M146	1.12745	2.42880
23 M148	2.44522	4.15960
24 M150	1.71473	3.52520
25 M159	.50509	1.20640
26 M164	1.47087	2.79000
27 M168	14.78174	38.12000
28 M169	4.57783	9.41160
29 M175	7.88304	25.69280
30 M182	8.76174	22.64560
31 M183	1.51913	4.87520
32 M192	40.39133	5.72160
33 M193	21.20120	4.57520
34 M196	4.82304	26.85120
35 M197	1.78391	7.37760
36 M200	1.00130	1.51280
37 M203	.58557	1.21160
38 M204	1.33503	3.75760
39 M205	1.01952	1.92880
40 M207	1.05913	2.09000
41 M208	4.24739	16.15100
42 M209	1.59078	4.93760
43 M210	2.57478	4.92480
44 M215	2.38217	4.92640
45 M224	1.76913	3.92880
46 M225	1.19913	2.11200
47 M228	1.19739	1.69200
48 M231	.54251	1.02760
49 M250	1.31952	2.42160
50 M318	1.35087	.35160
51 M319	.39685	.23760
52 NUM	1.00000	2.00000
OCCURTS	23.	25.

FIGURE 15

STEP NUMBER 1
VARIABLE ENTERED 35 M197

VARIABLE	F TO REMOVE	FORCE LEVEL
	DF= 1 45	
35 M197	77.320	1

VARIABLE	F TO ENTER	FORCE LEVEL	TOLERANCE
	DF= 1 45		
1 M74	19.711	1	.828849
2 M78	23.680	1	.834929
3 M79	37.924	1	.789451
4 M80	23.301	1	.838190
5 M86	12.253	1	.970530
6 M93	13.014	1	.944935

STEP NUMBER 2
VARIABLE ENTERED 8 M95

VARIABLE	F TO REMOVE	FORCE LEVEL
	DF= 1 45	
8 M95	60.325	1
35 M197	97.112	1

VARIABLE	F TO ENTER	FORCE LEVEL	TOLERANCE
	DF= 1 44		
1 M74	.092	1	.564376
2 M78	.485	1	.615665
3 M79	3.932	1	.630895
4 M83	.591	1	.635779
5 M86	.323	1	.655327
6 M93	1.099	1	.379373
7 M94	6.318	1	.389350
9 M96	3.561	1	.373784
10 M102	7.859	1	.955730
11 M103	12.946	1	.956764
12 M109	.098	1	.654129
13 M112	7.464	1	.964165
14 M116	11.512	1	.943255

STEP NUMBER 15
VARIABLE ENTERED 47 M223

VARIABLE	F TO REMOVE	FORCE LEVEL
	DF= 1 32	
6 M93	4.245	1
8 M95	5.950	1
11 M103	13.027	1
14 M115	26.043	1
15 M117	1.553	1
23 M143	24.362	1
24 M150	3.373	1
28 M169	7.913	1
30 M182	5.727	1
31 M183	15.431	1
32 M192	15.323	1
33 M193	14.922	1
35 M197	23.259	1
47 M223	1.568	1
50 M318	20.178	1

VARIABLE	F TO ENTER	FORCE LEVEL	TOLERANCE
	DF= 1 31		
1 M74	.528	1	.323941
2 M78	.341	1	.253890
3 M79	.316	1	.130986
4 M83	.955	1	.224236
5 M86	.096	1	.487871
7 M94	.387	1	.208220
9 M96	.079	1	.151766
10 M102	.275	1	.188740
12 M109	.039	1	.274432
13 M112	1.575	1	.527261
16 M118	.572	1	.558126
17 M121	.142	1	.393252
18 M136	.232	1	.329842
19 M140	.374	1	.556038
20 M142	.485	1	.524500
21 M145	.030	1	.352794
22 M146	.001	1	.422966
25 M159	.353	1	.275788
26 M164	.051	1	.582123
27 M168	.200	1	.375410
29 M175	.471	1	.409093
34 M196	.253	1	.347733
36 M203	5.400	1	.394898
37 M203	2.837	1	.358089
38 M204	.011	1	.354050
39 M205	.069	1	.527908
40 M207	.073	1	.471941
41 M208	.591	1	.435052
42 M209	.002	1	.287341
43 M210	.192	1	.441451
44 M218	.593	1	.331959
45 M224	.088	1	.565408
46 M225	.415	1	.382767
48 M231	2.079	1	.329413
49 M250	2.597	1	.414738
51 M319	1.075	1	.392925

U-STATISTIC OR WILKS' LAMBDA .0309221 DEGREES OF FREEDOM 15 1 46
APPROXIMATE F-STATISTIC 65.832 DEGREES OF FREEDOM 15.00 32.00

FIGURE 16

STEP NUMBER 40
VARIABLE ENTERED 45 M221

	VARIABLE	F TO REMOVE DF= 1 17	FORCE LEVEL		VARIABLE	F TO ENTER DF= 1 16	FORCE LEVEL	TOLERANCE
2	M78	3.954	1	1	M74	.238	1	.193106
3	M79	1.729	1	4	M85	.025	1	.134387
7	M94	32.721	1	5	M86	.002	1	.156815
10	M112	5.577	1	6	M93	.066	1	.042416
11	M113	50.119	1	8	M95	.001	1	.120466
15	M117	12.361	1	9	M96	.293	1	.031415
16	M118	3.662	1	12	M109	.315	1	.130397
18	M135	3.941	1	13	M112	.013	1	.257956
19	M140	25.389	1	14	M116	.046	1	.088561
20	M142	9.267	1	17	M121	.131	1	.120778
23	M148	12.870	1	21	M145	.190	1	.085782
24	M150	5.801	1	22	M146	.382	1	.117529
25	M159	7.320	1	26	M164	.284	1	.287812
28	M169	136.930	1	27	M168	.284	1	.064119
29	M175	2.253	1	30	M182	.003	1	.107116
31	M183	34.226	1	33	M204	.213	1	.089356
32	M192	11.075	1	34	M205	.143	1	.068208
33	M193	16.380	1	42	M209	.139	1	.017539
34	M195	6.565	1	44	M218	.004	1	.053936
35	M197	103.600	1	46	M225	.108	1	.101219
36	M200	46.629	1	51	M319	.247	1	.160053
37	M203	30.651	1					
40	M207	56.551	1					
41	M208	33.423	1					
43	M210	17.779	1					
45	M221	1.330	1					
47	M229	56.929	1					
48	M231	14.579	1					
49	M250	18.763	1					
50	M319	74.090	1					

SU-STATISTIC OF WILKS' LAMBDA .0023294 DEGREES OF FREEDOM 30 1 46
APPROXIMATE F-STATISTIC 292.701 DEGREES OF FREEDOM 30.00 17.00

FIGURE 17

OGROUP L

CASE				
1	31.5	1.000	1665.3	0.000
2	32.5	1.000	1629.6	0.000
3	15.9	1.000	1518.7	0.000
4	29.1	1.000	1910.9	0.000
5	35.7	1.000	1654.6	0.000
6	31.2	1.000	1484.1	0.000
7	37.5	1.000	1711.8	0.000
8	37.6	1.000	1603.4	0.000
9	25.8	1.000	1667.8	0.000
10	33.6	1.000	1710.2	0.000
11	22.2	1.000	1745.5	0.000
12	15.9	1.000	1536.8	0.000
13	17.1	1.000	1656.4	0.000
14	21.8	1.000	1721.7	0.000
15	23.4	1.000	1642.3	0.000
16	21.0	1.000	1605.8	0.000
17	13.4	1.000	1764.1	0.000
18	29.8	1.000	1616.1	0.000
19	25.9	1.000	1633.8	0.000
20	18.5	1.000	1737.9	0.000
21	29.3	1.000	1765.8	0.000
22	25.7	1.000	1673.6	0.000
23	28.0	1.000	1656.5	0.000

OGROUP C

CASE				
24	1707.1	0.000	38.4	1.000
25	1587.8	0.000	32.6	1.000
26	1825.0	0.000	30.2	1.000
27	1518.5	0.000	27.9	1.000
28	1755.4	0.000	39.4	1.000
29	1607.8	0.000	31.1	1.000
30	1675.4	0.000	28.5	1.000
31	1621.5	0.000	24.3	1.000
32	1691.7	0.000	25.6	1.000
33	1715.8	0.000	35.0	1.000
34	1692.4	0.000	28.7	1.000
35	1681.2	0.000	28.0	1.000
36	1529.2	0.000	28.5	1.000
37	1812.0	0.000	26.9	1.000
38	1732.4	0.000	26.3	1.000
39	1561.0	0.000	31.0	1.000
40	1661.5	0.000	28.3	1.000
41	1732.5	0.000	36.0	1.000
42	1682.9	0.000	23.5	1.000
43	1751.8	0.000	18.8	1.000
44	1612.4	0.000	33.4	1.000
45	1727.2	0.000	33.5	1.000
46	1675.2	0.000	35.4	1.000
47	1685.8	0.000	37.5	1.000
48	1619.0	0.000	42.6	1.000

POINTS TO BE PLOTTED

OVARIABLE		OGROUP	MEAN COORDINATES		SYMBOL FOR CASES	SYMBOL FOR MEAN
2	V78					
3	V79	L	21.12	.00	L	1
7	V83	C	-19.43	.00	C	2
10	V102					
11	V103					
15	V117					
16	V118					
18	V136					
19	V140					
20	V142					
23	V148					
24	V150					
25	V139					
28	V159					
29	V175					
31	V193					
32	V192					
33	V195					
34	V196					
35	V197					
36	V200					
37	V203					
40	V207					
41	V208					
43	V210					
45	V224					
47	V228					
48	V231					
49	V230					
50	V318					

OGROUP L

CASE	X	Y	CASE	X	Y	CASE	X	Y
1	20.99	.85	11	22.09	.69	21	21.51	1.72
2	20.56	-.80	12	19.50	-1.10	22	21.16	1.06
3	19.40	.19	13	21.57	-.41	23	21.42	-1.22
4	24.11	.38	14	21.80	-3.37			
5	20.83	.19	15	20.80	-.59			
6	18.76	-.46	16	20.38	1.56			
7	21.49	-.40	17	22.36	.65			
8	21.14	1.31	18	20.46	-.57			
9	21.08	-.65	19	20.69	-.71			
10	21.57	.83	20	22.04	1.06			

OGROUP C

CASE	X	Y	CASE	X	Y	CASE	X	Y
24	-19.73	-.87	34	-19.57	-.40	44	-16.62	-1.22
25	-18.32	-.18	35	-19.56	-.43	45	-20.04	.33
26	-21.28	.75	36	-19.52	-.06	46	-19.38	-.61
27	-17.76	-.33	37	-21.16	-1.04	47	-19.49	1.09
28	-20.44	1.90	38	-20.19	-.13	48	-18.95	-.50
29	-18.59	-.72	39	-18.01	-.26			
30	-18.47	2.04	40	-19.17	-.32			
31	-18.85	-.17	41	-20.07	.12			
32	-18.87	.13	42	-19.00	1.16			
33	-20.11	-.78	43	-20.52	.56			

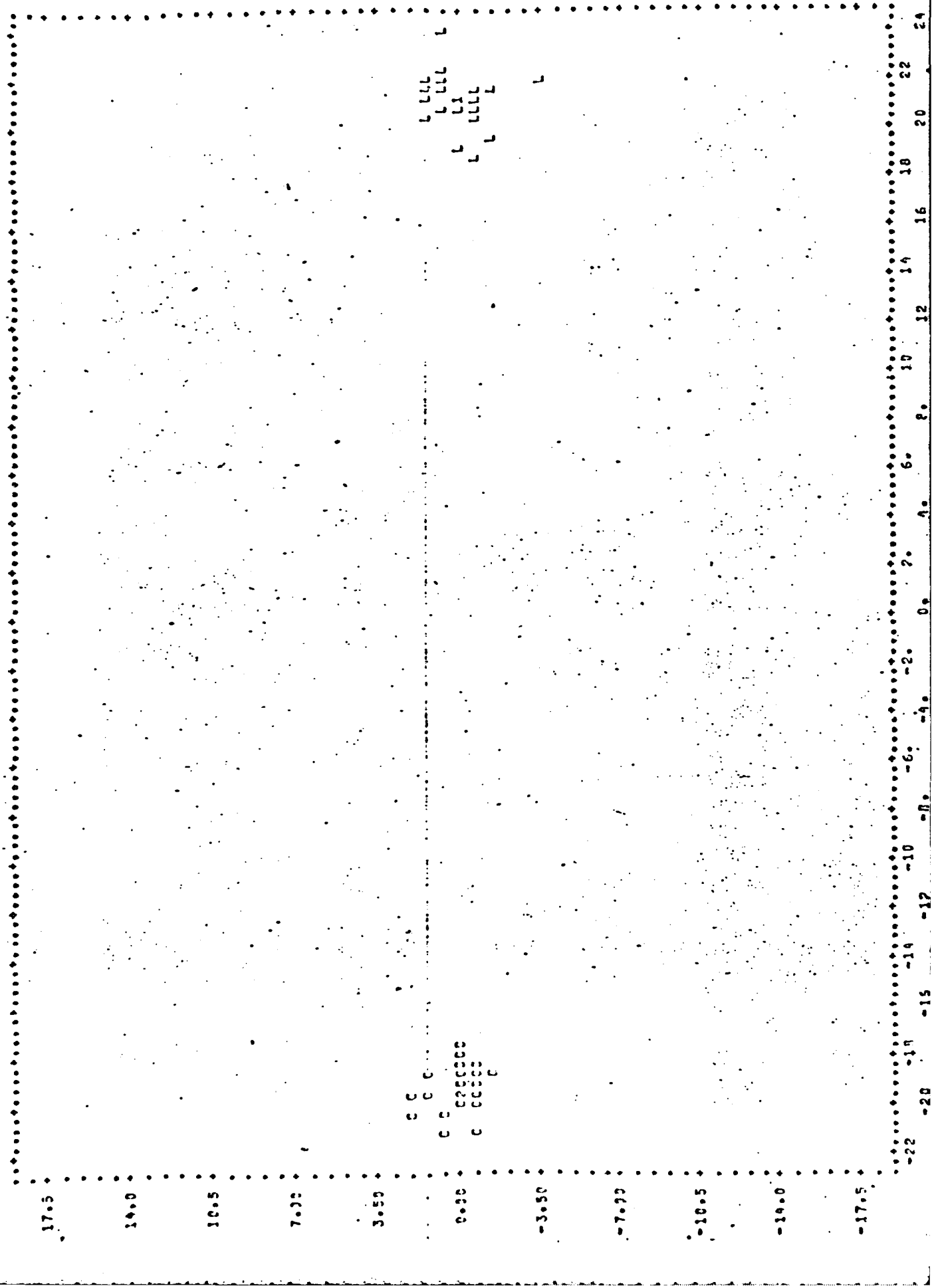


FIGURE 18

We see that the stepwise discriminant analysis has separated the cases in a more "decisive" manner than by the WNI test (Figure 11 vs Figure 18). This is not surprising in view of the fact that it was given the "best" preselected 50 variables and its own procedure used only 30 out of these for the ultimate classification. Although the quantitative and graphical presentations by this procedure are the most "convincing" among the three types of classification, it is the clustering analysis that demonstrated in an utterly unbiased fashion that in this study we had just two biochemically distinct groups.

The Wilcoxon-WNI program was recently modified to classify unknown samples. First two known groups are compared. "P-values" for each m/e and average spectra for each group are calculated. Second WNI's are calculated for each unknown sample from the average spectra of the two given groups. The weighting factor used was $P^{-1/2}$. This modification of the Wilcoxon-WNI program provides unbiased classification of unknown samples because they do not contribute to the calculations of the "p-values" and the average spectra of given groups, and it gives a direct comparison with the classification of unknown samples by discriminant analysis.

As stated elsewhere in this report we shall proceed with additional comparative statistical analyses of other series of biological samples, including sets less distinctly different than the one shown here, before deciding which statistical treatment is most suitable for a given problem.

D. EXPERIMENTAL RESULTS

1. Hospital Patient Studies

The entire multicomponent mixture analysis procedure has been tested using four sets of pathological and three sets of control samples. Two sets of samples representing liver malfunction have been obtained. One set of eight samples includes a variety of diseases, including malignancies, with liver involvement as a primary or secondary reason for hospitalization. Another set

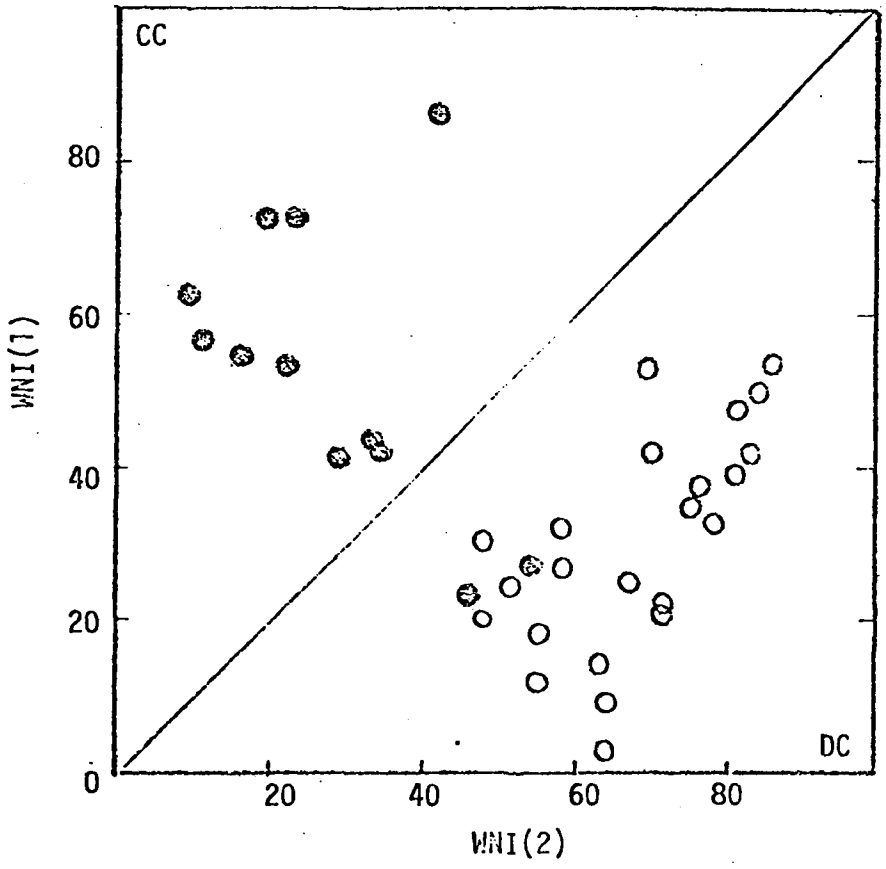
of 12 samples is more homogeneous representing primarily alcoholic hepatitis. A sample set from 7 individuals hospitalized in the same ward with the alcoholic hepatitis patients was evaluated as a control set. The reason for studying of patients with liver disorders was to obtain data on patients with liver involvement other than viral hepatitis, which has been studied by us earlier in this program (Clin. Chem. 22, 1503 (1976)).

Samples were provided by the staff of Childrens Hospital from 7 pneumonia cases (probably of viral origin), 12 diarrhea patients diagnosed as of viral origin (Rotavirus), and 6 samples selected from patients hospitalized for concussion and tonsilectomy, situations unlikely to involve an infectious organism. An additional set of 13 adult control samples was obtained from healthy volunteers within the university.

Each sample was analyzed in duplicate, interspersing control and pathological samples to avoid systematic errors. Summed spectra were assembled into data sets for processing in the CYBER by either the Wilcoxon-WNI programs or the BMDP programs.

The results of the Wilcoxon-WNI programs are presented for several data sets in figures 19 through 25. The graphs display WNI(1) versus WNI(2) for each sample in both groups. The WNI(1) axis represents the difference of a given sample from the average spectra of the pathological group while the WNI(2) axis shows the difference of a spectra from the control group class average. In this representation, control samples should fall close to the vertical axis and a maximum distance from the horizontal, while pathological samples should fall near the horizontal axis and away from the vertical axis. The comparison of the diarrhea in children with children controls (Fig. 19) shows two control samples falling within the pathological group, when all masses are used with a weighting function of 1/p to compute the WNI's. The

FIGURE 19



WNI PLOT OF CHILDREN'S CONTROL SAMPLES (CC-CLOSED CIRCLES) VS DIARRHEA IN CHILDREN (DC-OPEN CIRCLES) USING ALL m/e VALUES WEIGHTED BY 1/p FOR COMPUTING WNI'S

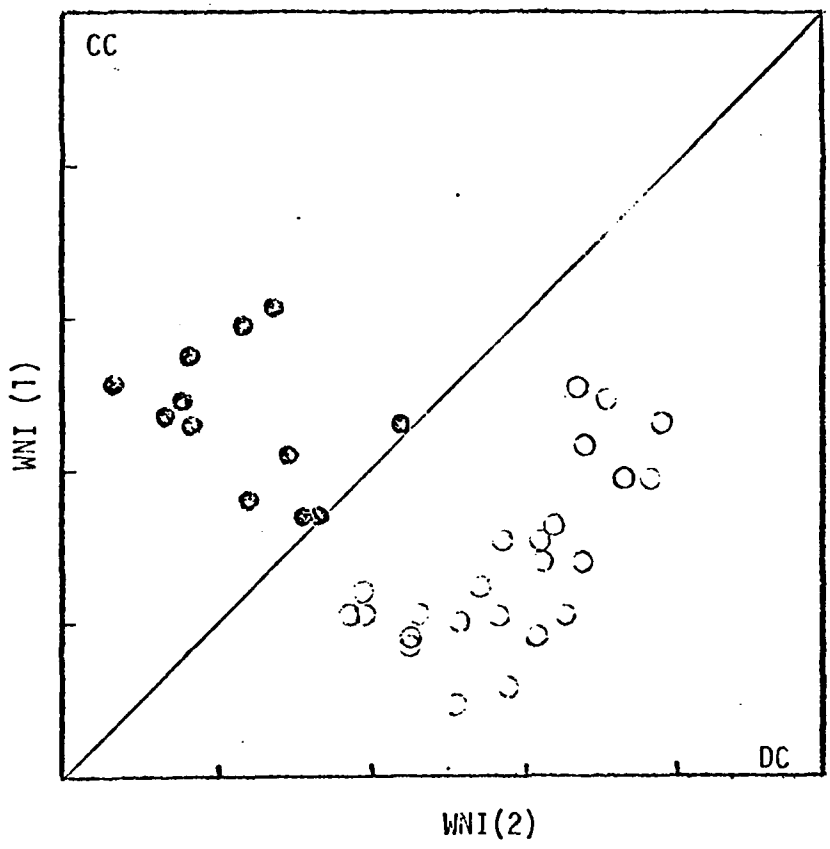
use of only the eight mass of lowest p-value (Fig. 20) not only shows complete separation of control and diarrhea cases, but seems to reveal a distinct grouping among the diarrhea cases that may be the result of a different response to the infecting organism, or perhaps due to a different etiology of the disease. For several of the diarrhea cases, the samples of whom were collected within a few days of each other, rotavivus was detected by independent tests.

The pneumonia cases show one individual from the pathological set falling within the control group using WNIs based on the 14 peaks of lowest p-value (Fig. 21). The hospital records for this child indicated that this patient was discharged the day after the sample was taken, so it might be surmised that this child had already recovered from the infection. Adding this individual to the control group and removing the sample pair from the pathological group did not appreciably change the clustering of samples obtained before, indicating that this sample properly belongs in the control group.

An example of differential diagnosis is given in Fig. 22 comparing the diarrhea to the pneumonia cases. Incomplete separation was obtained using all masses weighted as 1/p in the WNI calculation.

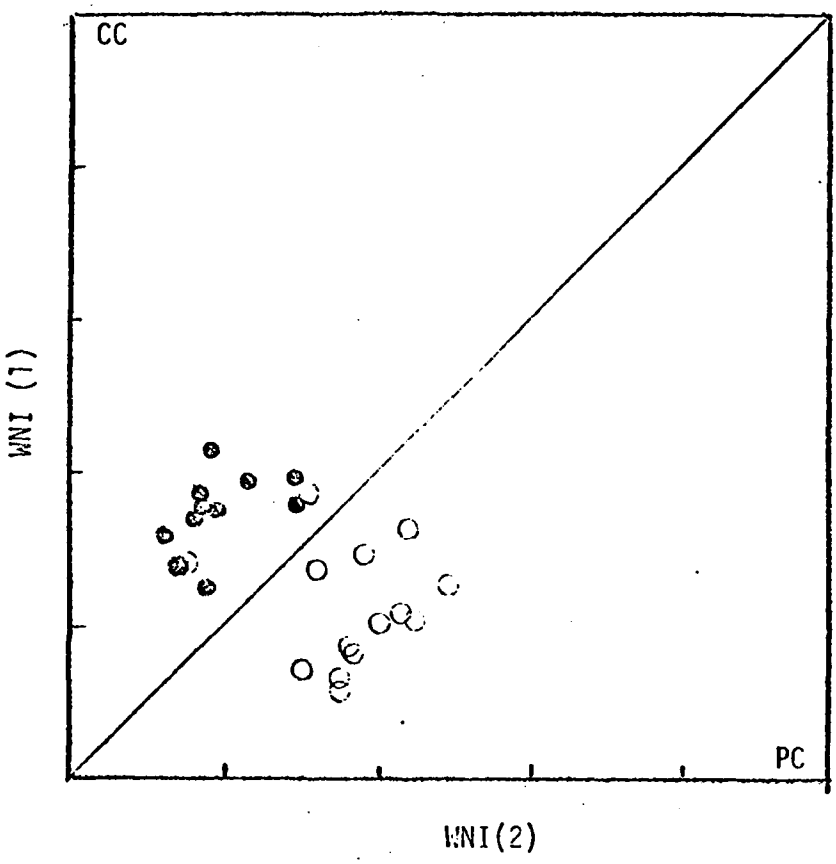
Figures 23 and 24 show the mixed set of liver disorders compared to healthy adult controls, using all masses weighted by 1/p and using only the 12 masses of lowest p-value. Again the use of only the diagnostic peaks improves the separation between the two classes. Comparing the alcoholic hepatitis set with hospital controls (Fig. 25) reveals a weakness in the choice of this control set, so that although the alcoholic hepatitis samples are tightly clustered, a number of samples from the control set fall within this grouping. In this experiment patients with gastric ulcer, functional gall bladder disease, and anemia appear to have similar profile patterns to those

FIGURE 20



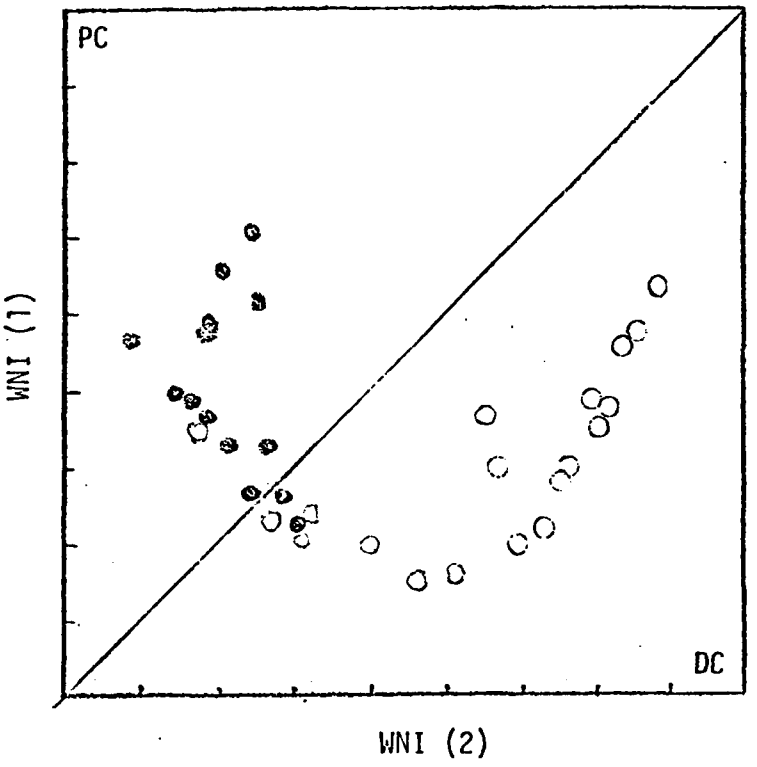
CONTROL (CC-CLOSED CIRCLES) VS DIARRHEA SAMPLES (DC-OPEN CIRCLES) USING 8 m/e VALUES OF LOWEST p VALUES TO COMPUTE WNI'S

FIGURE 21



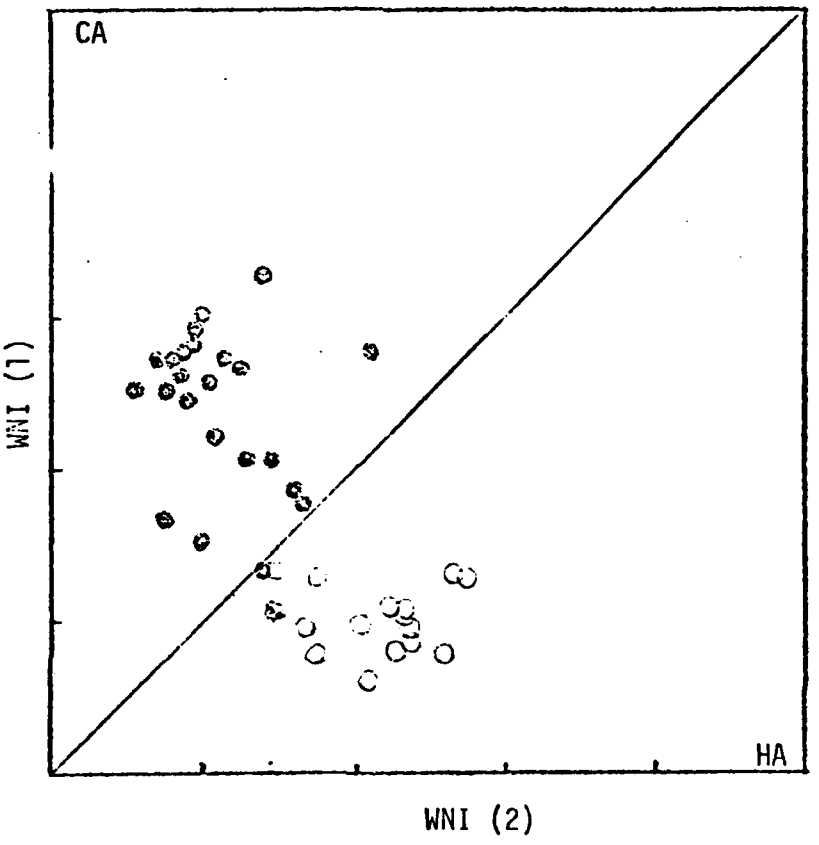
CHILDREN'S CONTROL SAMPLES (CC-CLOSED CIRCLES) VS CHILDREN'S PNEUMONIA SAMPLES USING 14 m/e VALUES FOR WNI'S (PC-OPEN CIRCLES).

FIGURE 22



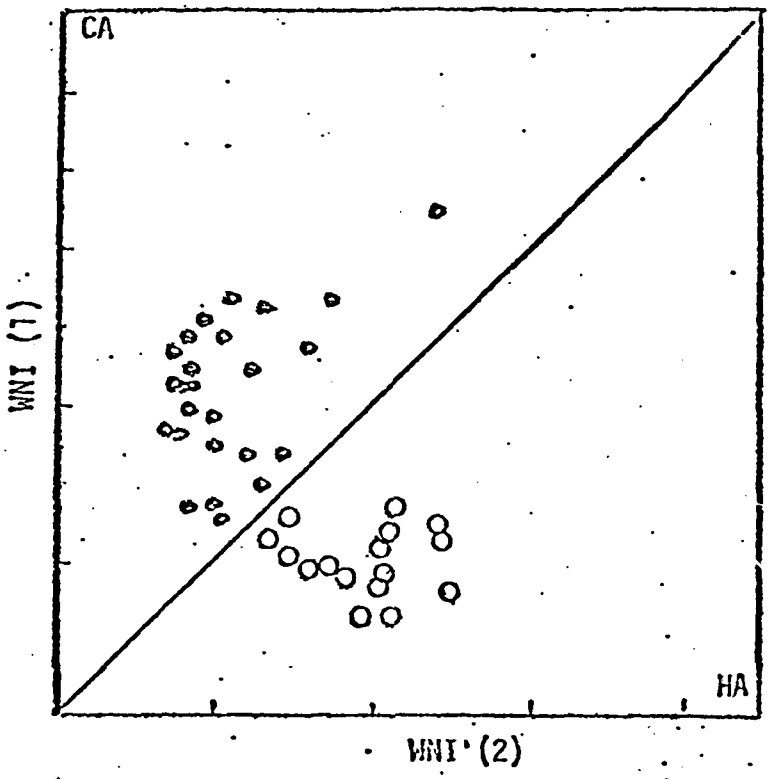
PNEUMONIA SAMPLES (PC-CLOSED CIRCLES) VS DIARRHEA SAMPLES (DC-OPEN CIRCLES)

FIGURE 23



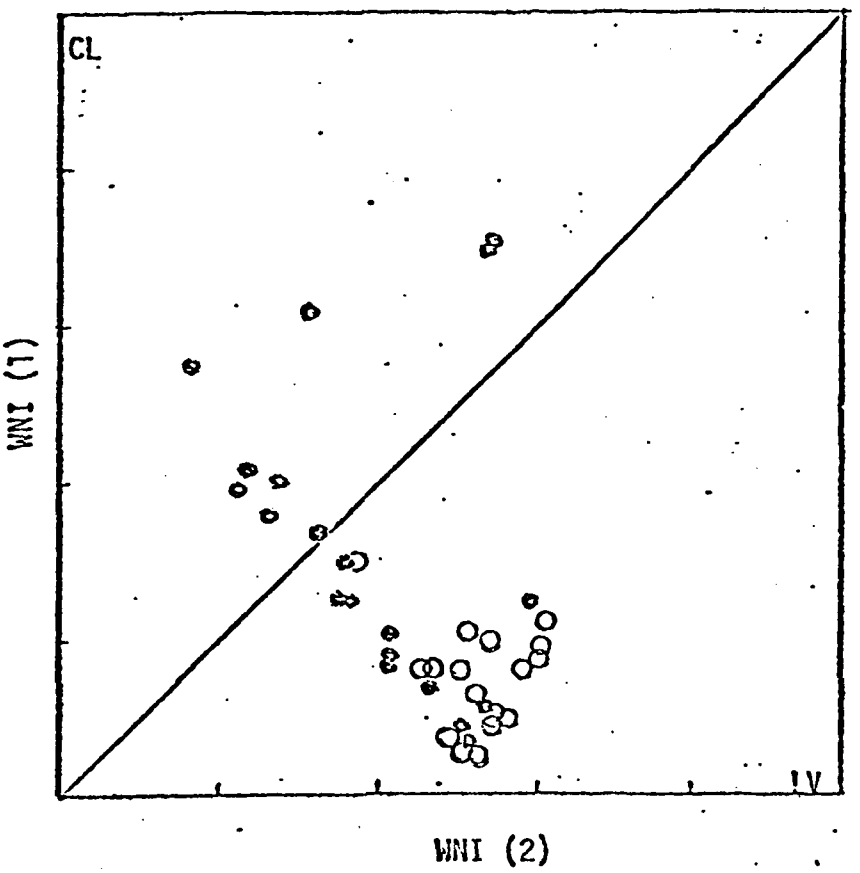
ADULT CONTROL SAMPLES (CA CLOSED CIRCLES) VS MIXED LIVER DISORDERS (HA-OPEN CIRCLES) USING ALL m/e VALUES.

FIGURE 24



ADULT CONTROL SAMPLES (CA-CLOSED CIRCLES) VS MIXED LIVER DISORDERS (HA-OPEN CIRCLES) USING 12 m/e VALUES.

FIGURE 25



HOSPITALIZED CONTROLS (CL-CLOSED CIRCLES) VS
ALCOHOLIC LIVER DISEASE (LV-OPEN CIRCLES)

with liver disorders. This example illustrates the importance of obtaining well characterized samples upon which to base the initial pattern analysis.

Comparing the alcoholic hepatitis set with the adult controls (Fig. 11) shows a much clearer separation. In both comparisons the set of pathological samples appears to be a well defined homogenous group. The use of the adult control set in the comparison generates, however, a different set of masses with the lowest p-values. In addition the p-values were 2 to 3 orders of magnitude smaller using the adult controls. Tabular results of the Wilcoxon-WNI analysis of these samples is presented in Figure 10.

These two groups of samples were also analyzed using the BMDP programs. First the BMDP 3D program was used to calculate the t-statistic, separate and pooled, for the peak intensities at each m/e value for the null hypothesis of equivalent means for both the pathological and control groups. Based on these results 51 masses were selected. A significant number of the masses selected by this method also show low p-values by the Wilcoxon test. Using these selected variables both cluster and stepwise discriminant analysis programs were used to classify the individual cases. These results have been described in section 3D.

2. Longitudinal Studies on Virus Infected Patients.

Controlled longitudinal studies were carried out on two sets of volunteers at the USAMRIID. One group consisted of seven individuals who received live virus vaccine for sandfly fever. Two additional individuals in this group received a placebo injection. None of the participants were told whether they received the vaccine or control injection. Morning urine samples were collected from paticipants 4 days prior to the injection, the day of the injection, for 8 consecutive days subsequent to injection and finally 28 days after the injection.

All samples were stored frozen without preservative, shipped to this laboratory packed in dry ice, and kept frozen at -20°C until analysis. We initially analyzed samples from three individuals in the sandfly fever experiment without any prior knowledge of the sample classification (vaccine vs. control) or the expected time and duration of the symptoms. The samples from each individual were prepared and analyzed in duplicate in a random sequence. In a few cases during this series, a sample had to be repeated due to instrument malfunction during data acquisition. In these cases we found that better replicate samples were obtined by using the salt saturated urine remaining from the previous extraction rather than using the original refrozen sample. One possible explanation is that a bacterial contamination may have affected the original samples during their exposure to room temperature. We did not observe similar differences in repeated analyses of samples that are collected and stored with $ZnSO_4$ as a preservative. However additional samples of the same urines obtained from USAMRIID at a later date and analyzed 9 months later showed substantial changes from the first set of samples, indicating that storage at -20°C, at least without a preservative to inhibit enzymatic activity, does not guarantee the preservation of sample composition.

The data from the first three individuals of the sandfly fever showed significant temporal differences in the patterns of two of the individuals (Joffe, LeBlanc) compared to the third (Berry).

Throughout the sampling period Berry's profiles show significantly lower intensity over the entire mass range. This individual also apparently did not consume any caffeinated beverages during the study, leading to a further qualitative difference in this person's pattern compared to the other two participants. We, therefore, did not include Berry's patterns in the initial analysis of the data. (We have received the information that this individual

served as one of the controls after the analysis of all samples were completed). With only two individuals to examine we were concerned that dietary variations plus individual variations in the response to the vaccine might obscure the detection of the pathological pattern or location of the maximum response period. Two approaches were partially successful in handling this problem of deciphering an unknown pattern appearing at an unknown time in a small data set.

In the first approach the BMDP 7D program for variable stratification was employed to show the means and intensity distributions for each m/e value as a function of the sample collection sequence. An example is shown in Figure 26. This analysis, although tedious and inefficient for routine use, did reveal a substantial number of masses showing changes in intensity distributions at days 3, 4 and 5. The second approach involved comparing "arbitrarily" patterns of days minus 4 and zero against days 3, 4, and 5. Days 3-5 were also the expected time for development of symptoms for the particular virus employed.

The t-test program P3D was used to select 40 m/e values showing maximum differences between the control period (days -4 and 0) and the expected response period (days 3 through 5). The clustering analysis program was applied to the two virus infected sample sets using the selected m/e values from the baseline days -4 and 0, day 2, and days 3 through 5, the latter presumably showing changes associated with the response to the virus infection. The results of this analysis are shown in Figure 27. It should be noted that this analysis involves no preselection of the number of groups needed to classify the samples. The results show the control days -4 and 0 in one cluster with the presumed virus response days, samples of day 3, 4, and 5 in a second cluster. The day 2 samples from one individual clustered with the

control days group of samples while corresponding samples from the other individual belonged to the group of samples of days 3 to 5.

The same 40 variables were used in the discriminant analysis program to derive a classification function for separating days -4 and 0 from days 3,4, and 5. This classification function was used to classify sample days 2,7,8 and 28 of the same individuals. These results are shown in Figure 28. The classification of day 2 samples was the same as obtained in the clustering analysis. We have been informed by USAMRIID that these individuals showed a similar difference in their rates of return to a normal profile following the response period. Neither subject developed symptoms prior to the middle of day 3 and both subjects were free of symptoms by day 7. The observed pattern differences in morning urine samples on day 3 thus preceded the appearance of clinical symptoms and persisted following their disappearance. The third set of samples from a control subject was classified as healthy throughout by the same classification procedure.

Unfortunately in this series no samples were collected beyond day 8 so that it is impossible to determine at this point the time of unambiguous disappearance of the pathological pattern. In any case this pattern in urine seems to appear prior to the clinical symptoms and to persist beyond the time of their disappearance.

The samples from the remaining participants in this study were shipped to us at a later date and days -4, 0, 3, 4 and 5 from persons receiving the vaccine were selected for initial analysis. During the preparation of this set the addition of EDTA was inadvertently omitted from the extraction procedure. Our statistical analysis of this sample set shows a detectable pattern difference due to this change in the procedure making a direct classification of all seven vaccinated individuals more complicated. The use

FIGURE 26

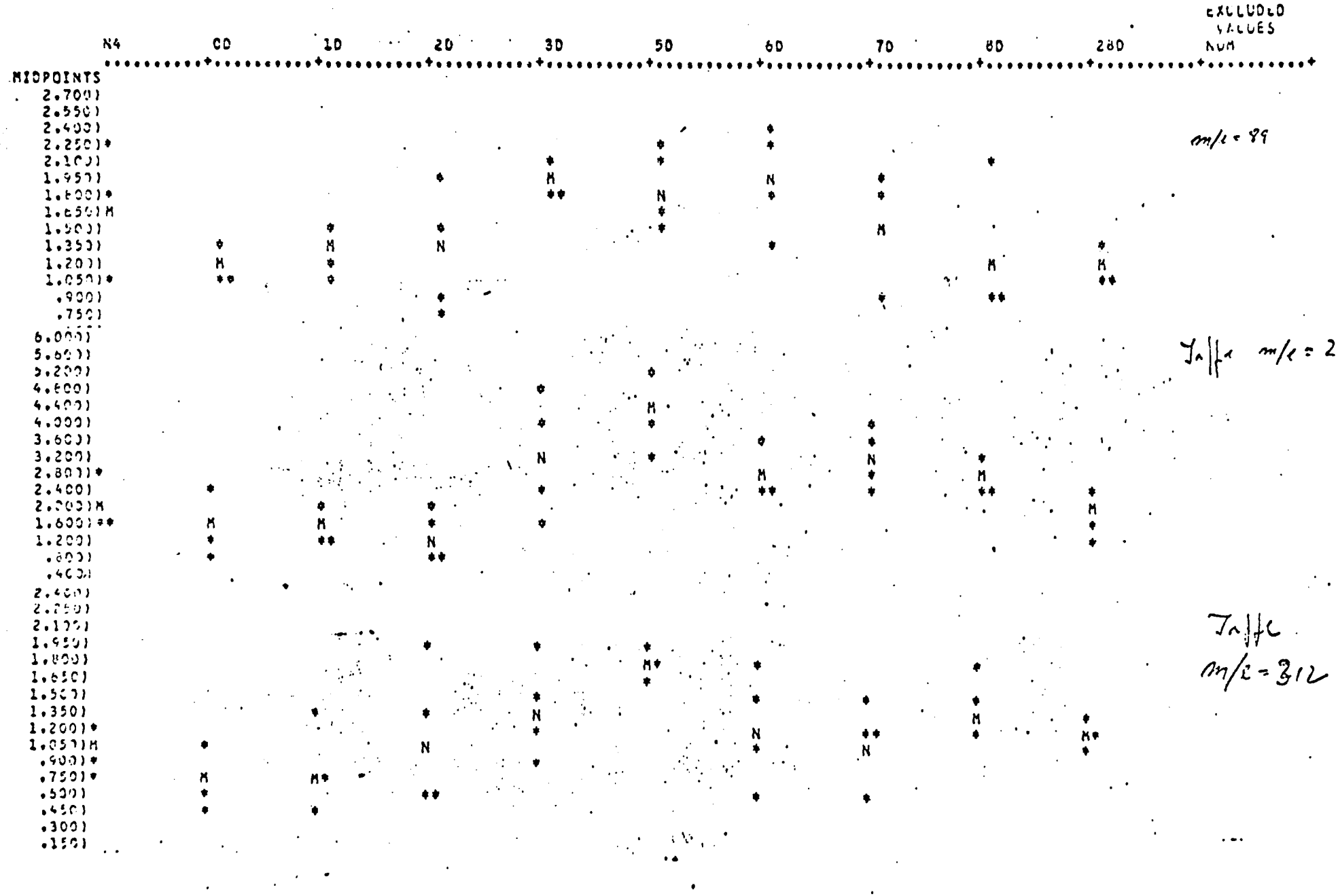

EXCLUDED VALUES
NUM
N4
CO
10
20
30
50
60
70
80
280
MIDPOINTS
m/e = 89
Jaffe m/e = 238
Jaffe
m/e = 312

FIGURE 27

CASE			ORDER OF AMALGAMATION
No	Day	Ind.	
1	4	J	****
5	4	L	6.
7	0	L	3.
8	0	L	2.
6	-4	L	4.
4	0	J	5.
2	-4	J	12.
3	0	J	13.
12	2	L*	1.
11	2	L*	16.
21	5	L	20.
20	4	L	19.
15	5	J	18.
18	3	L	17.
14	3	J	15.
16	5	J	11.
13	3	J	14.
10	2	J*	10.
19	4	L	8.
22	5	L	7.
17	3	L	9.
9	2	J*	21.

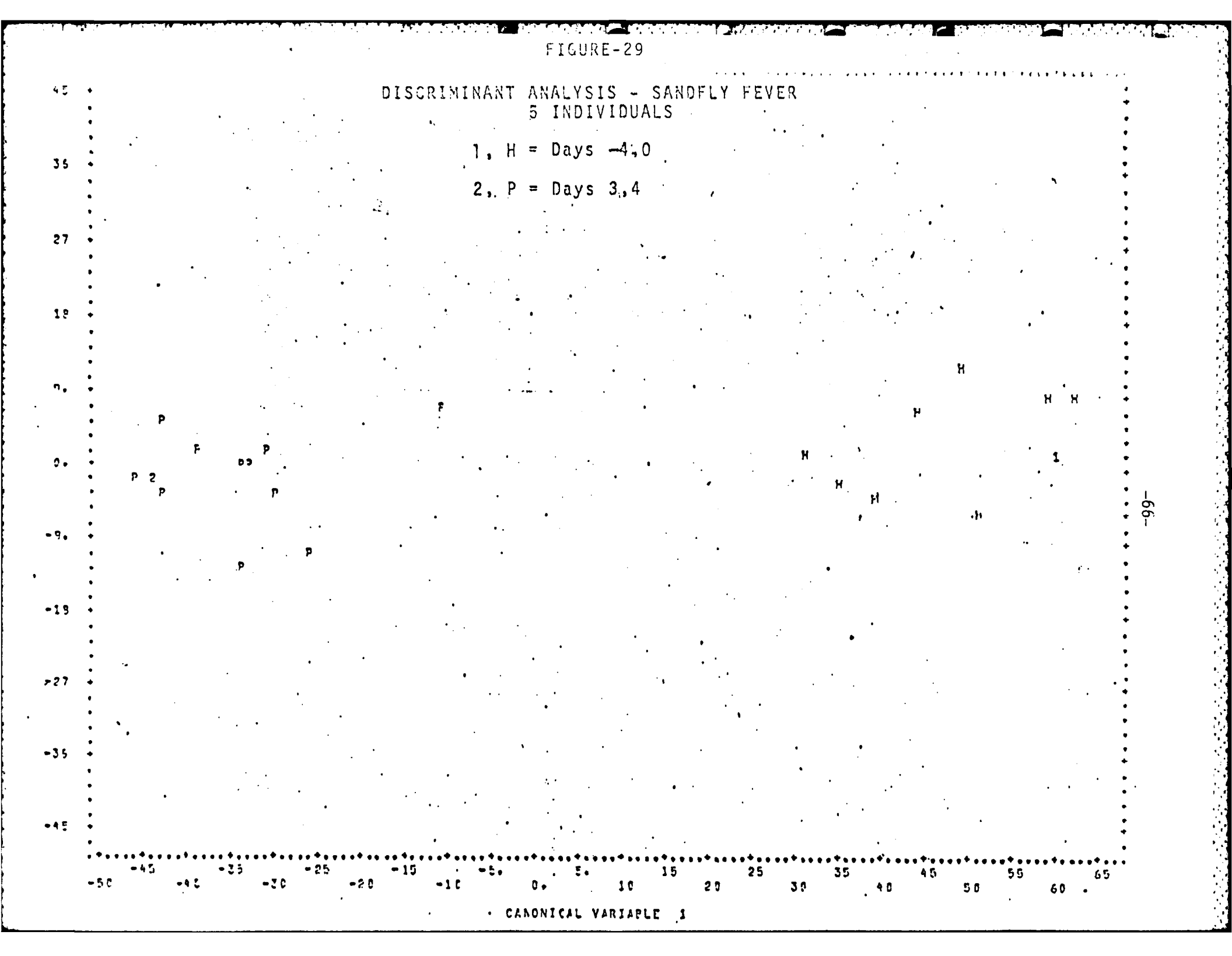
FIGURE 28

DISCRIMINANT ANALYSIS SANDFLY FEVER

	INCORRECT CLASSIFICATIONS	MAHALANOBIS D-SQUARE FROM AND POSTERIOR PROBABILITY FOR GROUP =		
GROUP A		A	C	
CASE				
1		14.8 1.000	****** 0.000	DAYS -4, 0; Joffe and LeBlanc
2		14.8 1.000	****** 0.000	
3		14.8 1.000	****** 0.000	
4		14.8 1.000	****** 0.000	
5		14.8 1.000	****** 0.000	
6		14.8 1.000	****** 0.000	
7		14.8 1.000	****** 0.000	
8		14.8 1.000	****** 0.000	
GROUP B		A	C	
CASE				
9	C	****** 0.000	****** 1.000	Cases 9,10 Day 2 Joffe
10	C	****** 0.000	****** 1.000	
11	A	****** 1.000	****** 0.000	Cases 11; 12, Day 2 LeBlanc
12	A	****** 1.000	****** 0.000	
GROUP C		A	C	
CASE				
13		****** 0.000	14.4 1.000	
14		****** 0.000	14.4 1.000	
15		****** 0.000	14.4 1.000	
16		****** 0.000	14.4 1.000	
17		****** 0.000	14.4 1.000	
18		****** 0.000	14.4 1.000	
19		****** 0.000	14.4 1.000	Days 3, 4, 5 Joffe, LeBlanc
20		****** 0.000	14.4 1.000	
21		****** 0.000	14.4 1.000	
22		****** 0.000	14.4 1.000	
GROUP D		A	C	
CASE				
23	A	****** 1.000	****** 0.000	Cases 23-25, Day 7,8,28 Joffe
24	A	****** 1.000	****** 0.000	
25	A	****** 1.000	****** 0.000	
26	C	****** 0.000	****** 1.000	Cases 26-28, Days 7,8,28 Joffe
27	C	****** 0.000	****** 1.000	
28	A	****** 1.000	****** 0.000	Cases 29-31, Days 7,8,28 LeBla
29	C	****** 0.000	****** 1.000	
30	C	****** 0.000	****** 1.000	Cases 32-34, Days 7,8,28 LeBla
31	A	****** 1.000	****** 0.000	
32	C	****** 0.000	****** 1.000	
33	C	****** 0.000	****** 1.000	
34	A	****** 1.000	****** 0.000	

of the t-test comparing normal and "pathological" periods of this last set of samples leads to a selection of variables that correctly classifies these samples using the discriminant analysis program (see Figure 29). The use of these variables with the discriminant analysis program to classify samples from all seven individuals into four groups was also successful. However, when we attempted to analyze the total sequence of samples of 2 individuals of the new set, together with the urine samples of one of the two individuals of the first set (LeBlanc) some 6 months later (9 months after the first set of samples was analyzed - in the interim we carried out the Cl study and the in vitro experiments described below), we were disappointed to find out that the composition of the samples has changed significantly. Although distinction between the normal and pathological days was still possible, the separation was not as distinct. The results are readily apparent looking at Figure 30. This figure shows a cluster analysis of the original LeBlanc samples (designated by L) and the LeBlanc samples that were analyzed 8 months later (designated by SFFL) using the 40 masses which were selected by the t-test of the old LeBlanc and Joffe samples of the control and pathological days. Looking at the "old" LeBlanc samples, one notices a definite separation between the control days (-4 and 0, designated L01 and LN41) and the pathological days (3 through 5, designated L3, L4, and L5). It can be seen that there are two clusters L41, L51, L31, and that of all other samples. The "new" LeBlanc samples (designed SFFN4L2, etc.) both control and pathological, form a cluster together with the control days of the "old" LeBlanc. One notices, however, that in that cluster there is still a definite difference between days 4 and 5 of the new LeBlanc (SFF4L2 and SFF5L2) and the rest of the samples in that cluster. However, day 3 (SFF3L2) clusters with the control samples.

FIGURE-29


DISCRIMINANT ANALYSIS - SANDFLY FEVER
5 INDIVIDUALS
1, H = Days -4,0
2, P = Days 3,4
45
35
27
18
9.
0.
-9.
-18
-27
-35
-45
-50
-45
-40
-35
-30
-25
-20
-15
-10
-5.
0.
5.
10
15
20
25
30
35
40
45
50
55
60
65
CANONICAL VARIABLE 1

Clearly the samples have changed with time. Of all the participants in the sandfly fever study, LeBlanc had the most fever hours and originally he had the strongest profile pattern. The other participants had less fever hours and one might not be surprised, therefore, that their patterns were not as distinct. The samples of some of the other participants (also analyzed 8 months after the original samples) gave similar results to the chemical ionization work - a less distinctive separation.

Since considerable care and effort is required in the FI methodology we also tested a simplified profiling procedure consisting of analysis of urease digested uring by isobutane chemical ionization mass spectrometry. The sample size of one microliter was empirically established as allowing a relatively rapid sample evaporation, while still maintaining a large excess in the isobutane reagent ion intensity. The longitudinal sample set of each individual were prepared as described in section B2, and analyzed in duplicate in a random sequence. For each sample set an aliquot of a "standard" urine from a healthy subject was incubated with enzyme at the same time. A sample of this standard urine, as well as a water blank containing the enzyme and buffer, were analyzed at the beginning of each sample set and a replicate of the standard sample was repeated at the completion of each set of samples of a given subject.

One objective of this methodology, a high sample throughput, was achieved in that the entire set of longtudinal samples from all nine participants, approximately 200 samples including blanks and standards, was completed in one week.

FIGURE 30

The patterns obtained were less informative compared with those of the extracted samples and analyzed by FI. This is not surprising since the sample size used limits the constituents examined by this procedure to species excreted at the levels above one mg per day. A representative spectrum shown in Figure 31 is dominated by the protonated molecular ions of creatinine (m/e 114) and hippuric acid (m/e 180). Direct examination of urine by field ionization using 50-75 microliters (Figure 32) gives an almost identical pattern except for the appearance of some constituents one amu lower as non-protonated molecular ions.

Analyses of the CIMS patterns by the same programs applied to the field ionization data yielded a rather weak separation. A t-test selection of variables and discriminant analyses using a subset of single replicate samples from all infected subjects resulted in a classification function that separated the control days (-4 and 0) from the fever days (4 and 5). Classifications of each person's sample series by this classification gave, however, 8Δ false positives and 27Δ false negatives. This result reflects the low signal to noise ratio of diagnostic information in this completely non-selective pattern of major urinary constituents which is not expected to be as significantly affected by the mild infection resulting from the experimental vaccination.

Although these findings suggest the inadequacy of the simplified sample preparation procedure for the detection of subtle metabolic changes associated with the infection, they indicate that at least in terms of the major urinary constituents, no major dietary or environmental differences have been represented in this sample set. Thus the more subtle pattern differences detected in the extracted samples are obtained in the context of a normal, unperturbed baseline pattern. However, another very important aspect has to

FIGURE 31

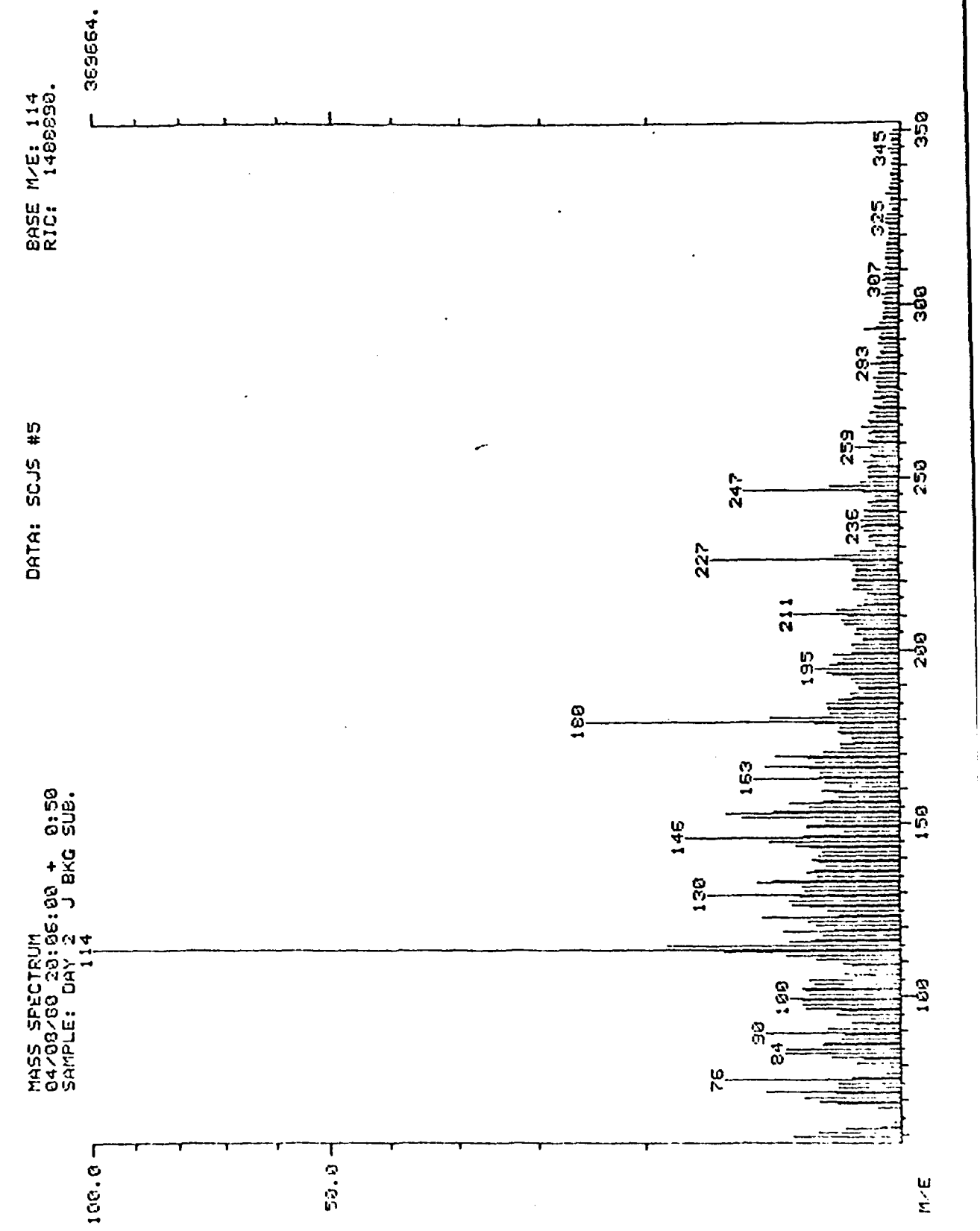
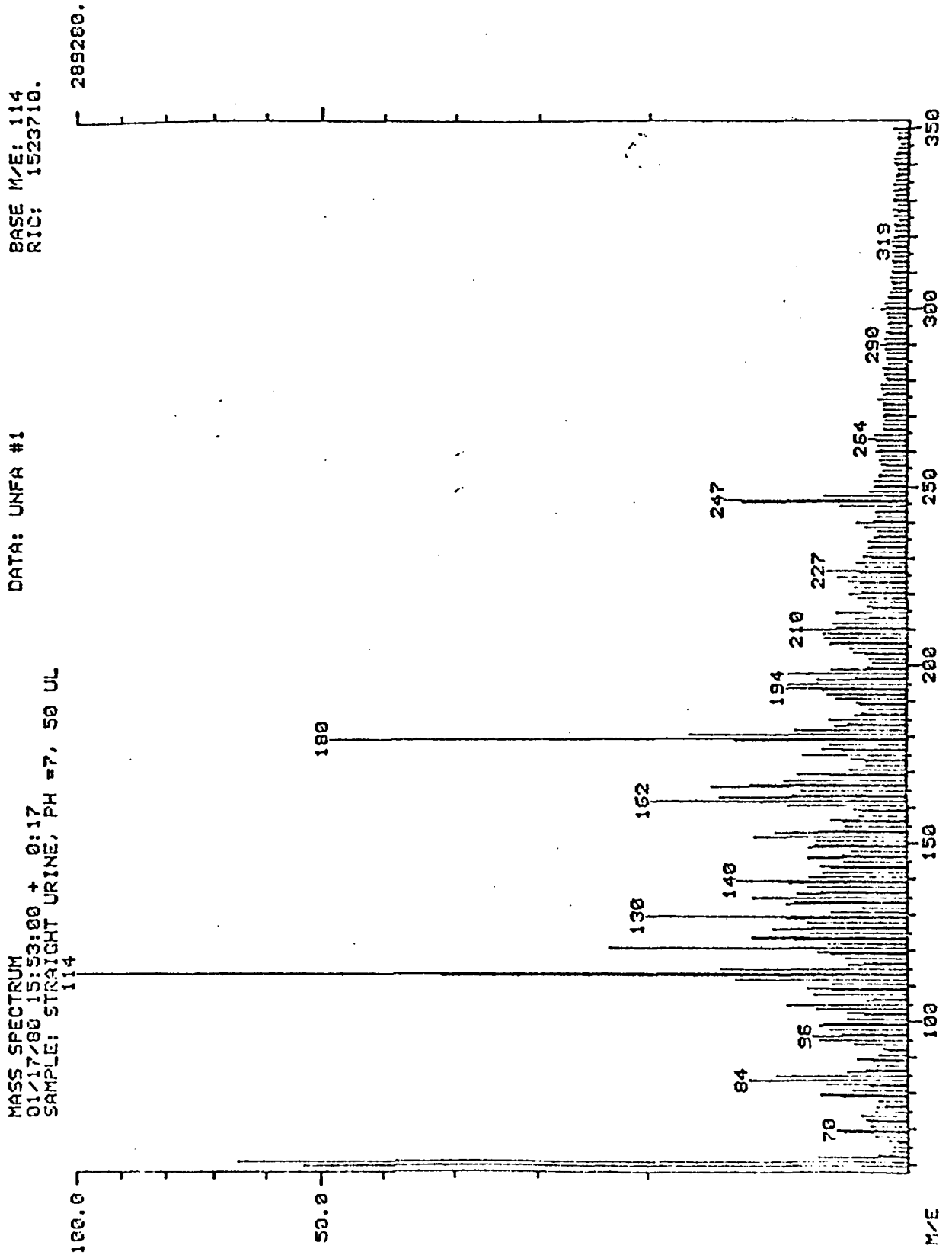

MASS SPECTRUM
04/08/80 20:06:00 + 0:50
SAMPLE: DAY 2 J BKG SUB.
DATA: SCJS #5
BASE M/E: 114
RIC: 1486890.
369664.
100.0
50.0
M/E
114
76
84
90
100
130
146
153
180
195
211
227
247
259
293
307
325
345
350

FIGURE 32


MASS SPECTRUM
01/17/80 15:53:00 + 0:17
SAMPLE: STRAIGHT URINE, PH =7, 50 UL
DATA: UNFA #1
BASE M/E: 114
RIC: 1523710.
289280.
100.0
50.0
M/E
114
180
247
319
70
84
96
130
140
162
194
210
227
264
290
300
350

be taken into consideration at this point. These samples, which were stored at -20°C, were approximately a year old when they were finally analyzed. Results of our earlier urine storage study indicated that samples stored at -20°C showed little changes in pattern after being kept for a month. Although at that time these changes were not significant, storage at -20°C for 10 or 12 months seems to be utterly inadequate. The repetition of FI analysis on the same samples subsequent to the CI analyses (see above), corroborated this conclusion. It remains to be seen if another longitudinal study on vaccinated individuals using appropriate preservation of storage would not be amenable to the simplified CI procedure.

3. Study of Human and Animal Tissue Cultures Infected With Polio Virus.

This study was undertaken to determine the differences between infected and uninfected cell cultures and to see how soon after infection of the cells such differences can be detected in the culture medium. The first study was performed with a cell line of human embryonic lung and the Mahoney strain of polio virus. The tissue culture work was carried out by Dr. Howard Faden and his associates at the Children's Hospital in Buffalo.

The human embryonic lung culture was originally established by Dr. Fishaut of Denver by 7 to 8 passes of tissue obtained from an aborted fetus. These cells were grown in Eagles minimal essential growth medium which was supplemented with 5Δ newborn calf serum. Cultures with equal numbers of cells were prepared in 5 ml plastic culture tubes that were incubated at 37°C for zero, 6, and 24 hours. For each of these three time points, nine tubes were prepared. Three tubes were left uninfected, three were infected with 1/10 ml of viral solution with a tissue culture infective dosage ($TCID_{50}$) of 1×10^2, and three tubes with a TCID of 1×10^4.

These titers of virus were allowed to adsorb onto the cells for one hour. At that time the cells were washed with isotonic phosphate buffer solution to remove excess virus particles. Following the wash, new medium was added to the tubes, and the cells were incubated for the times indicated. When the incubation times were reached, the tubes were refrigerated and centrifuged to separate the cells. We received the chilled, cell free, supernate of this centrifugation.

To insure inactivation of the virus, 20 microliters of .3M $HgCl_2$ was added to each tube. To one ml of this medium we added .25 ml of cold 70Δ perchloric acid and .25 ml of cold 11 M KOH to precipitate the proteins. The supernate of each tube was removed and diluted with an equal volume of distilled water to facilitate pH adjustment. This diluted medium was then titrated with KOH and $HClO_4$ to a pH of 2, 7, or 10, respectively. A 150 microliter sample was removed at each pH and placed in a glass culture tube which contained a folded 3 cm x 1 mm strip of Whatman fiberglass paper GP/C. The samples were evaporated onto this paper by blowing warm, dry nitrogen over them. The paper, with the sample dried onto it, was then introduced into the mass spectrometrer.

Initially high and zero virus samples incubated for zero and 24 hours from each of the 3 tubes were prepared at pH 2, 7, and 10 in duplicate. The data were first analyzed using the t-test to find peaks showing differences at 24 hours between zero and high virus. Using masses selected in this manner, cluster analysis and discriminant analysis programs were used to classify the four groups consisting of 24 hours high and zero virus and 0 hours high and zero virus. A better separation among the four groups was observed for the samples prepared at pH 2 (see Figure 33 through 35). Based on this analysis pH 2 was selected for preparation of high and low virus samples from 6 hours and 24 hours.

FIGURE 33

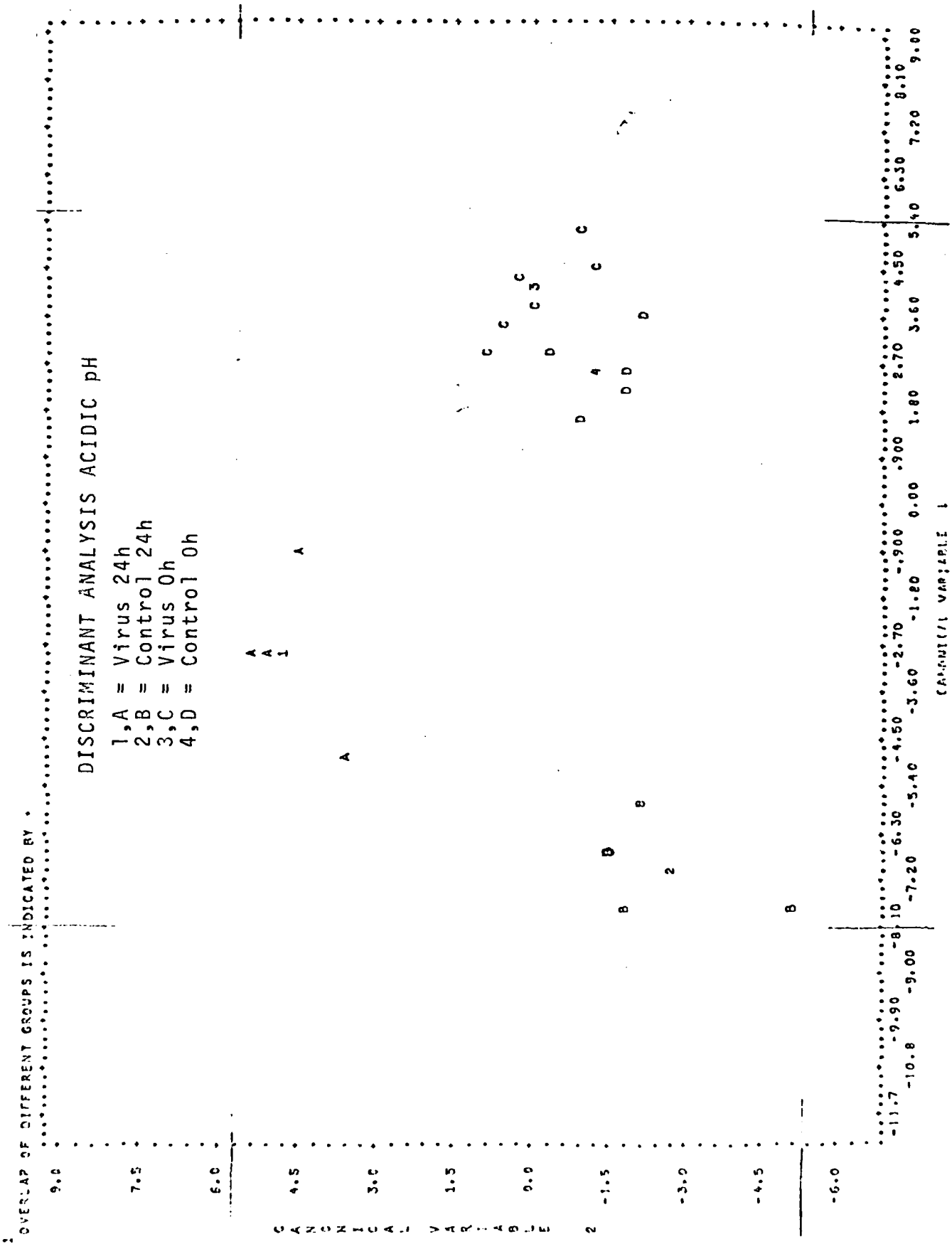

DISCRIMINANT ANALYSIS ACIDIC pH
1,A = Virus 24h
2,B = Control 24h
3,C = Virus 0h
4,D = Control 0h
OVERLAP OF DIFFERENT GROUPS IS INDICATED BY *
CANONICAL VARIABLE 2
CANONICAL VARIABLE 1

FIGURE 34

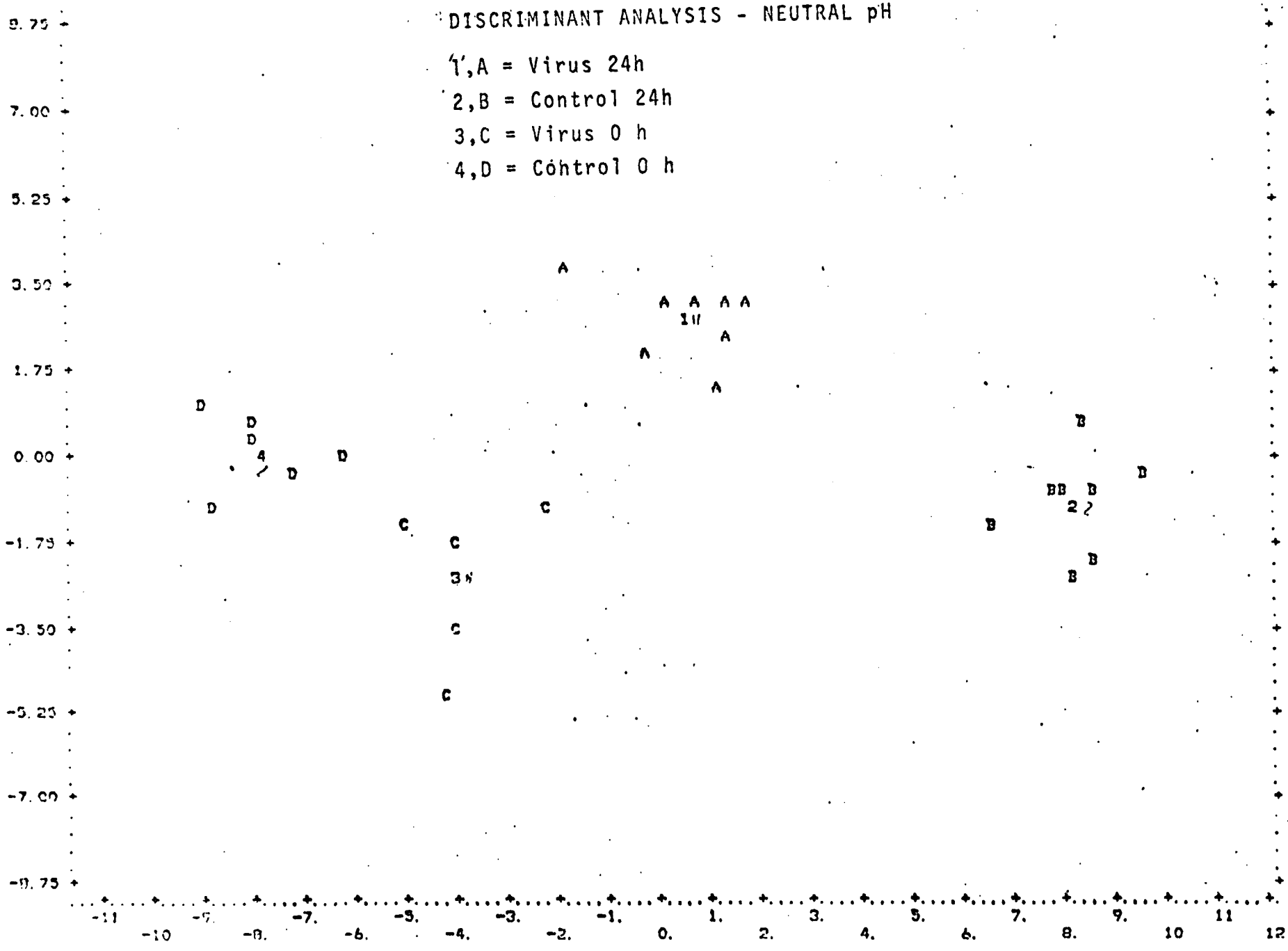
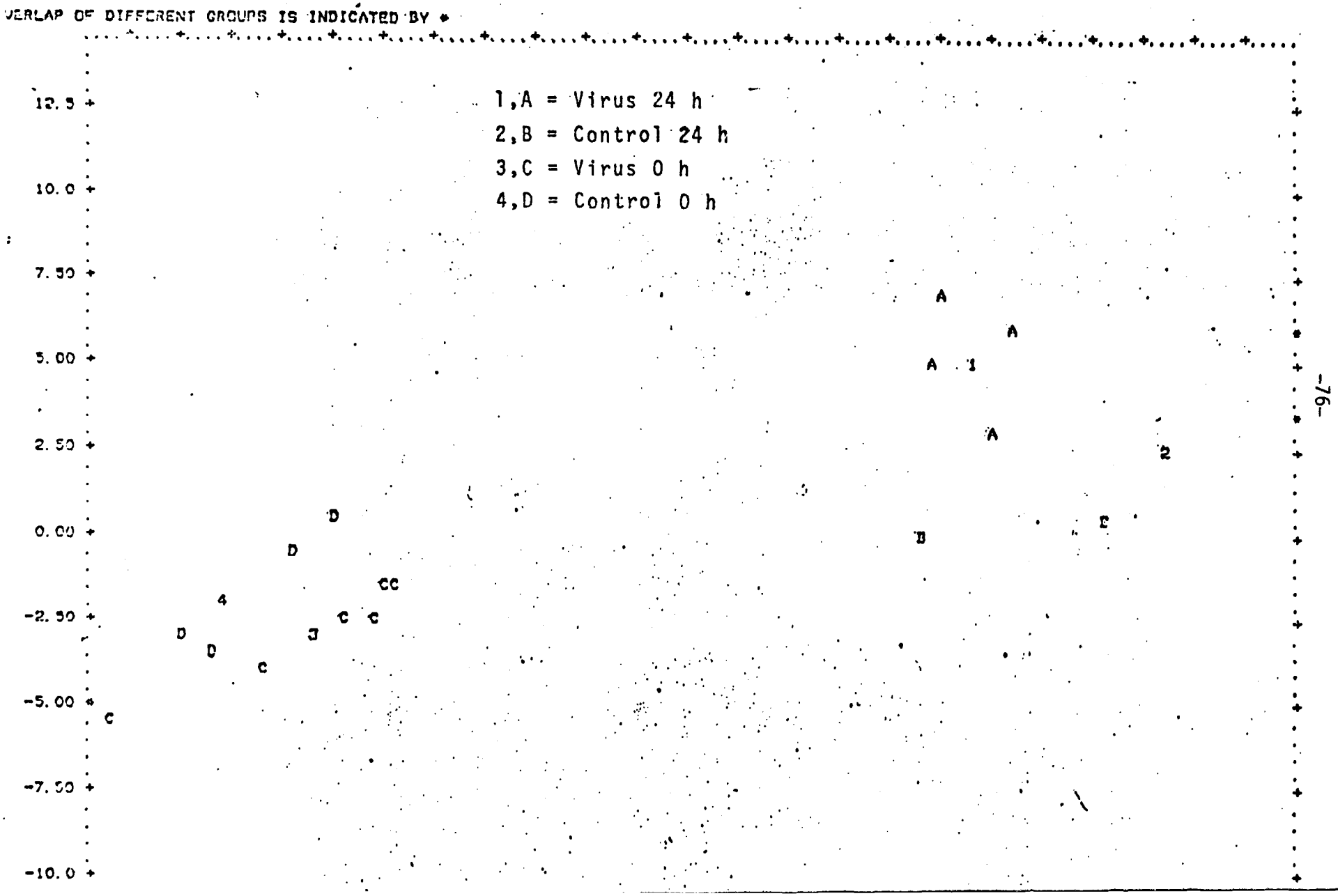

DISCRIMINANT ANALYSIS - NEUTRAL pH
1,A = Virus 24h
2,B = Control 24h
3,C = Virus 0 h
4,D = Control 0 h
9.75
7.00
5.25
3.50
1.75
0.00
-1.75
-3.50
-5.25
-7.00
-8.75
-11
-10
-9
-8.
-7.
-6.
-5.
-4.
-3.
-2.
-1.
0.
1.
2.
3.
4.
5.
6.
7.
8.
9.
10
11
12

FIGURE 35

DISCRIMINANT ANALYSIS - BASIC pH



The initial classification into four groups revealed several trends in the patterns. First there was a small but consistent difference in the zero time patterns of different virus concentrations, indicating a change in the composition of the media due to the adding of virus. This is an artifact not normally considered in cell culture experiments and it can be eliminated in future studies by making all virus innoculations including the zero virus using a common medium. Since that experiment we have tried to overcome this artifact by using UV light to deactivate the virus in the so called "virus free" samples and use deactivated virus medium as dilutant in the control experiments. We have shown that UV irradiation sufficient to deactivate the virus completely does not introduce significant changes in the mass spectrometric profile.

The incubation time and cell growth leads to time dependent changes in the intensity of selected m/e values. We observed both increases, which may be due to cell metabolites excreted into the media, and decreases possibly reflecting depletion of nutrients in the media. The third detectable trend is for the virus infected tubes to show similar time dependent trends but of reduced magnitude, consistent with an inhibition in the effective cell growth rate. That is, the same peaks which increase or decrease in uninfected cultures show a smaller increase or decrease respectivly in the infected cultures. Finally and more interestingly, there are a limited number of m/e's showing intensity changes that may specifically reflect virus activity. These include cases where virus infected cultures show larger changes in the same direction observed in non-infected tubes.

Two additional t-tests were used to obtain new groups of m/e values. By performing the t-test on zero versus 24 hours, at high virus concentration, a set of peaks was obtained reflecting both time and virus dependent pattern changes.

Similarly, a set of zero time media composition dependent peaks was found using the t-test to compare high and zero virus at zero time. Three new sets of variables were generated from the initial 42 variables selected using the t-test comparison between zero and 24 hours. For each set of variables the discriminant analysis program was used to develop a classification function for five groups consisting of the following:

group 1: all zero hour samples

group 2: high virus, 6 hrs

group 3: high virus, 24 hrs

group 4: zero virus, 6 hrs

group 5: zero virus, 24 hours.

A set of 27 masses was selected from the above 42 by removing 15 masses that also appeared in the zero hour test group.

The results are summarized in the canonical variable plots of the group average locations in Figures 36 through 38. The separation obtained with 27 variables (Figure 36) was increased when two additional zero time dependent masses were removed as shown in Figure 37. Reducing these 25 variable to 21 by eliminating 4 additional peaks in common with the zero time difference lead to reduced separation as shown in Figure 38. Thus it is not possible to eliminate all zero time artifacts without losing some of the time and virus specific information contained in the variables removed in the third test. This compromise, however, will hopefully not exist in future experiments designed to avoid initial differences in media composition.

These preliminary results on the "metabolic profile" of tissue culture media indicates that the virus infection can be detected in vitro 6 hours following exposure. We have continued our in vitro study of tissue cultures by first improving on the sample handling procedure (in addition to removal of the zero time artifact by UV irradiation as described above).

FIGURE 36

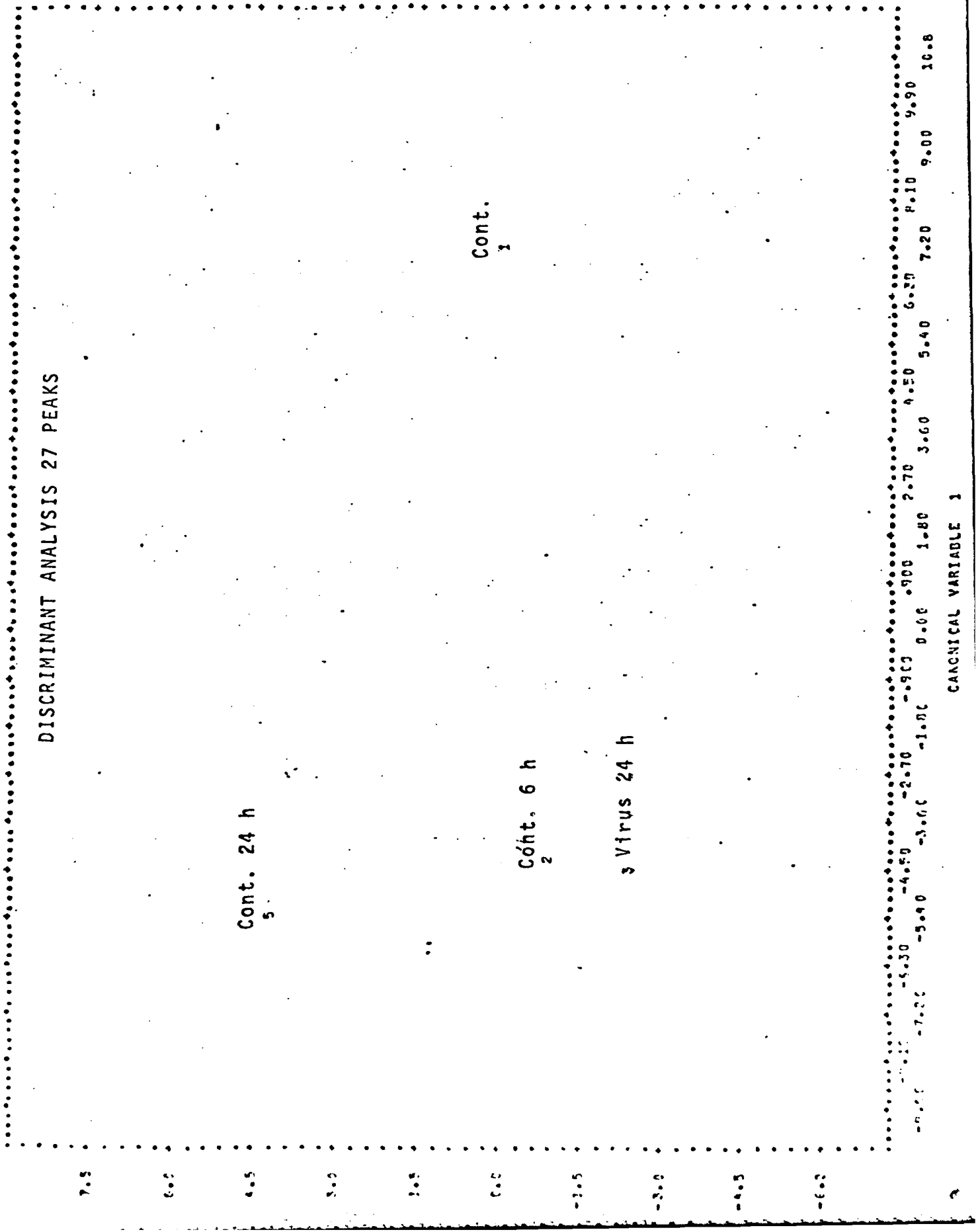

DISCRIMINANT ANALYSIS 27 PEAKS
Cont.
1
Cont. 24 h
5
Cóht. 6 h
2
s Virus 24 h
3
CANONICAL VARIABLE 1

FIGURE 37

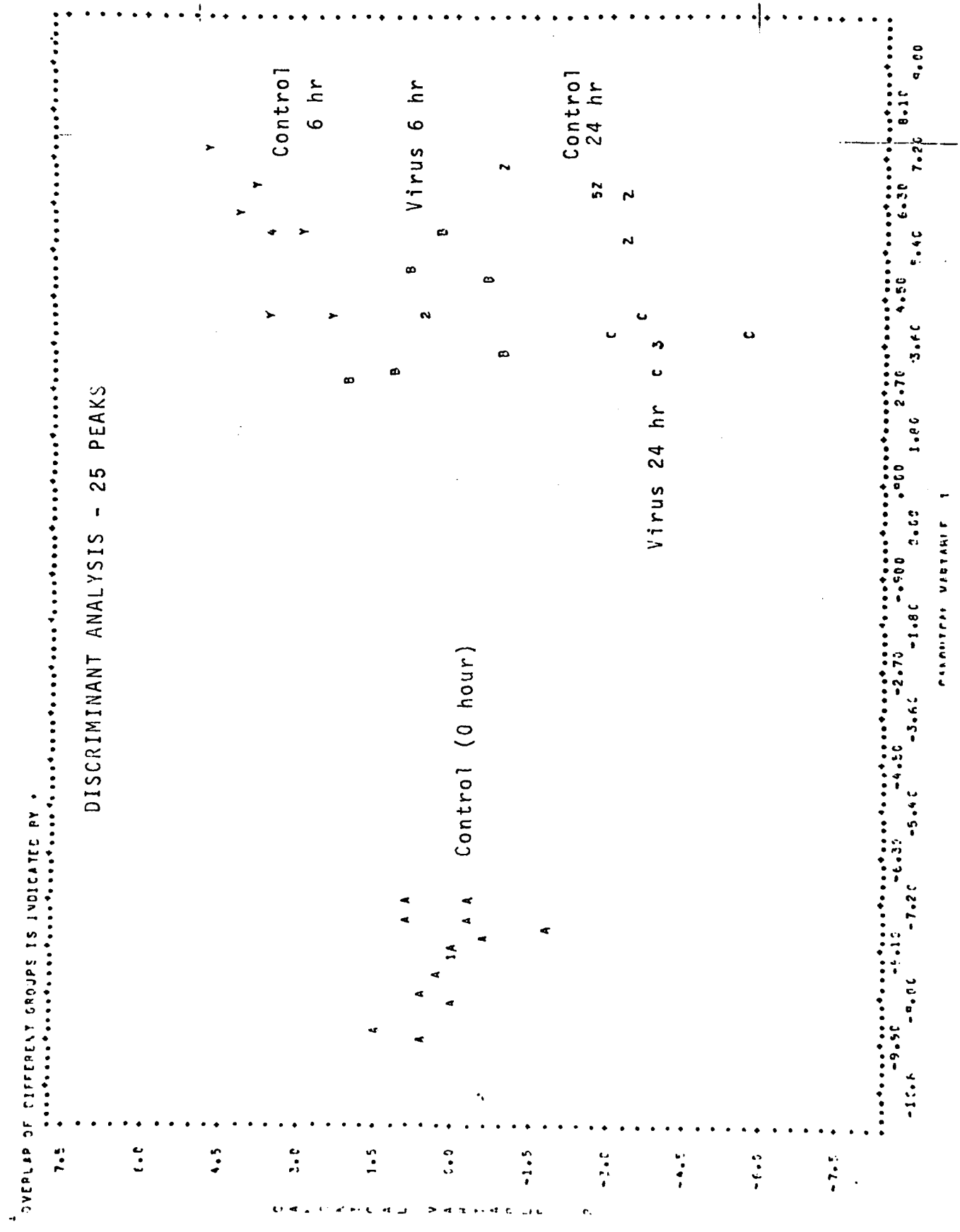
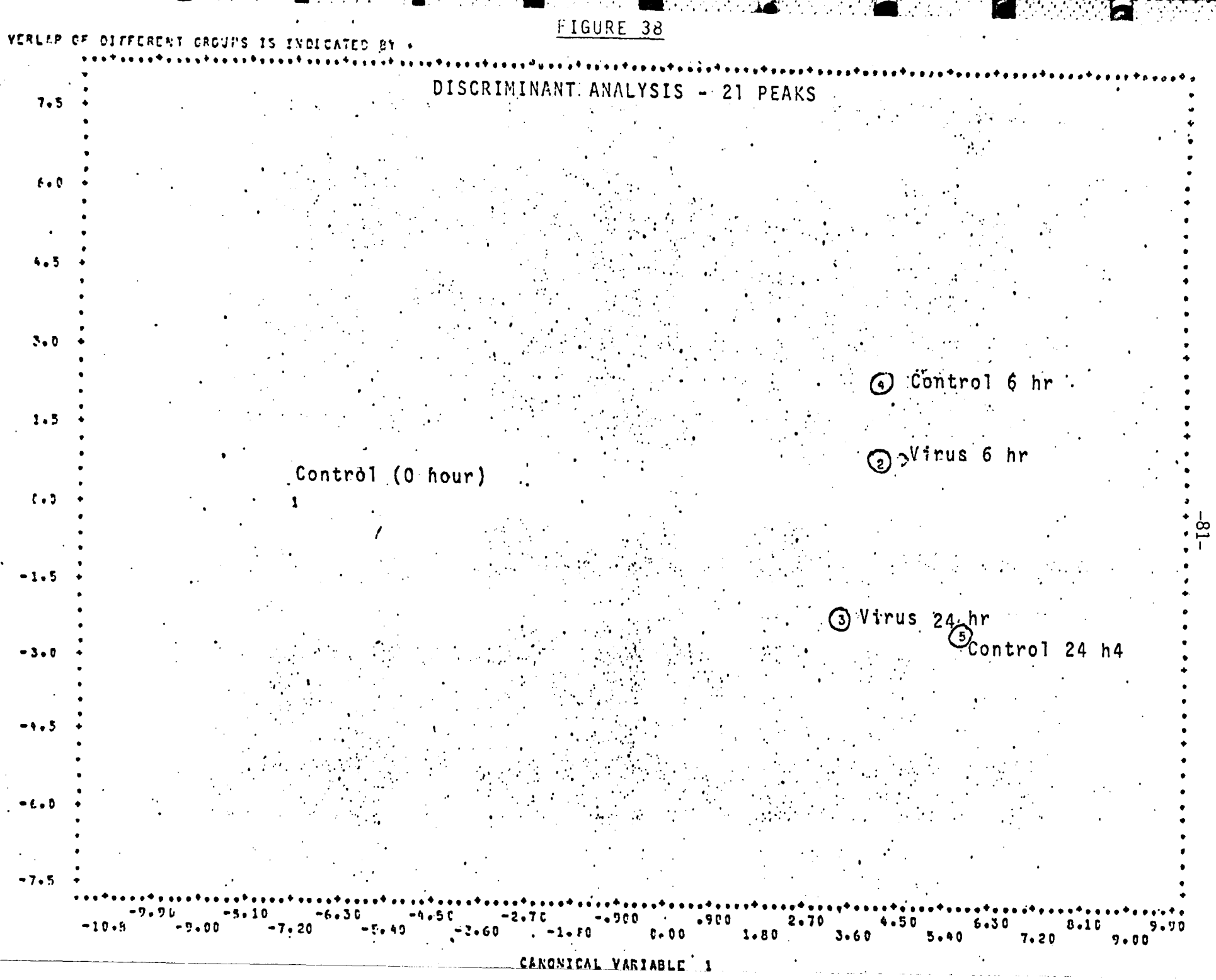

DISCRIMINANT ANALYSIS - 25 PEAKS
Control (0 hour)
Control 6 hr
Virus 6 hr
Control 24 hr
Virus 24 hr
OVERLAP OF DIFFERENT GROUPS IS INDICATED BY *

FIGURE 38


VERLAP OF DIFFERENT GROUPS IS INDICATED BY .
DISCRIMINANT ANALYSIS - 21 PEAKS
7.5
6.0
4.5
3.0
1.5
0.0
-1.5
-3.0
-4.5
-6.0
-7.5
-10.8
-9.90
-9.00
-8.10
-7.20
-6.30
-5.40
-4.50
-3.60
-2.70
-1.80
-.900
0.00
.900
1.80
2.70
3.60
4.50
5.40
6.30
7.20
8.10
9.00
9.90
CANONICAL VARIABLE 1
Control (0 hour)
1
4 Control 6 hr
2 Virus 6 hr
3 Virus 24 hr
5 Control 24 h4

At this stage we tested several different methods to eliminate protein from the tissue culture media, including precipitation with 70Δ perchloric acid, precipitation with ethanol, and filtration using Millipore immersible-CX ultrafilters (10,000 molecular weight cut-off). Comparing the mass spectra of the same media using the different pre-separation techniques, it was found that protein remove was not necessary, since the protein seems to have little effect on the profile pattern. In fact, the results (Figure 39-41) indicate that certain metabolites are lost through co-precipitation with the protein. Therefore it was decided to analyze the ext series of tissue culture samples without removing the protein. We have also optimized the pH of the mass spectrometrically analyzed samples to pH 2.2 where minimal variances due to imprecise pH adjustment do occur.

In our next study virus infected cultures were prepared by adding 0.1 ml of medium containing a tissue culture infective dose ($TCID_{50}$) of 10^3 or 10^4 virus particles/cell of Mahoney strain of polio virus. After allowing one hour for virus adsorption to cells, the infected cultures were washed with isotonic phosphate buffer (5 mmol/L) to remove excess virus particles and the tubes were refilled with fresh medium. Control samples were pre-incubated with a UV inactivated aliquot of the virus culture used for infected samples. A cell line of human embryonic lung tissue was cultured in Eagle's minimal essential growth medium supplemented with 5Δ newborn calf serum. Cultures with equal numbers of cells, prepared in 5 ml plastic culture tubes, were incubated at 37^0C. Three tubes each were prepared for analysis at zero, 6, 12, 24, and 48 hours for both infected and control cultures. Cell cultures were innoculated with a solution consisting of 0, 10, or 100Δ of the growth medium containing active virus particles. At the selected incubation times tubes were refrigerated (4^0C) and then centrifuged (1000 x g) to obtain the cell-free medium for analysis.

FIGURE 39

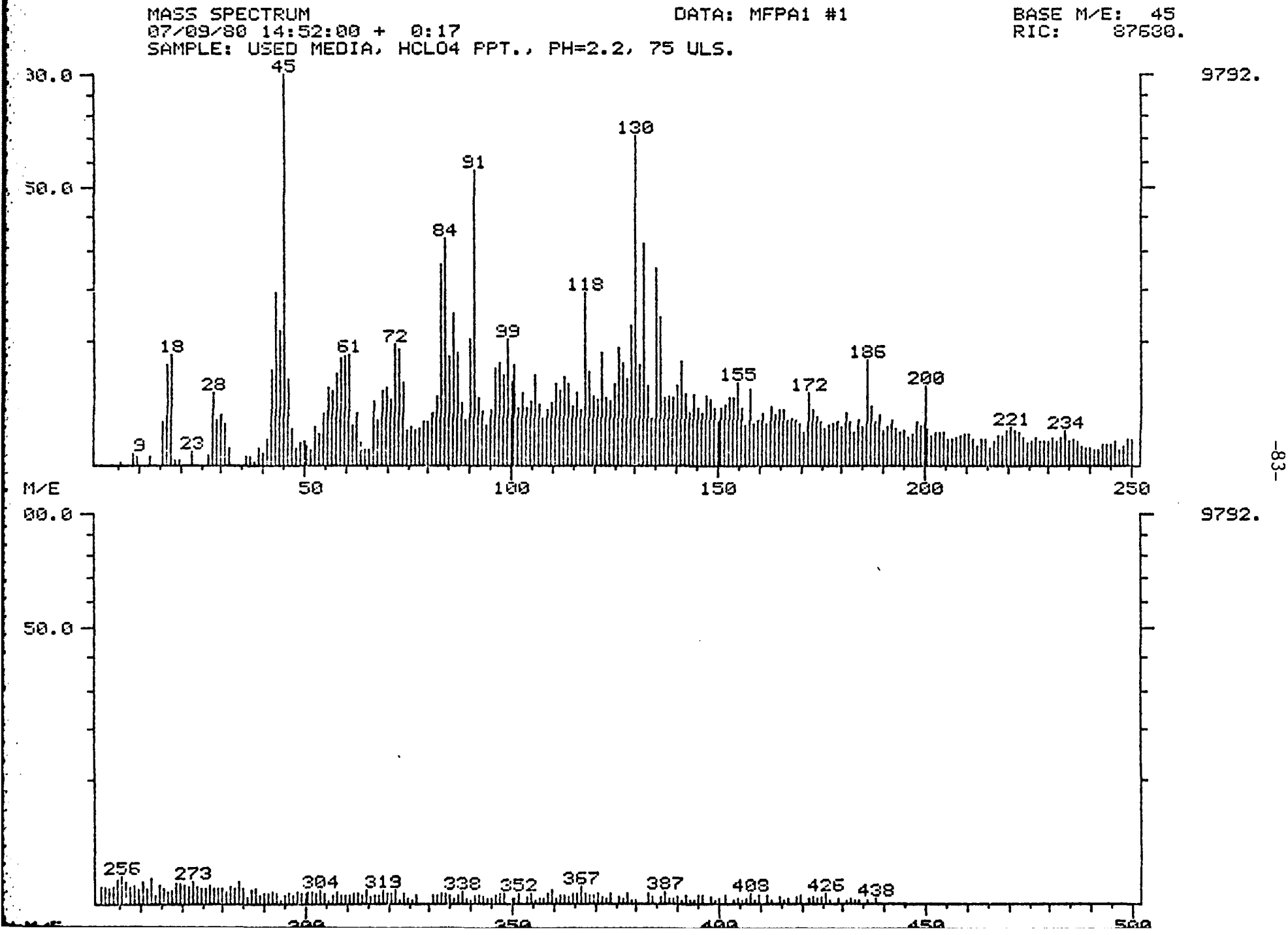

MASS SPECTRUM
07/09/80 14:52:00 + 0:17
SAMPLE: USED MEDIA, HCLO4 PPT., PH=2.2, 75 ULS.
DATA: MFPA1 #1
BASE M/E: 45
RIC: 87680.
100.0
50.0
9792.
45
130
91
84
118
99
72
61
18
28
9
23
155
172
186
200
221
234
M/E
50
100
150
200
250
100.0
50.0
9792.
256
273
304
313
338
352
367
387
408
426
438

FIGURE 40

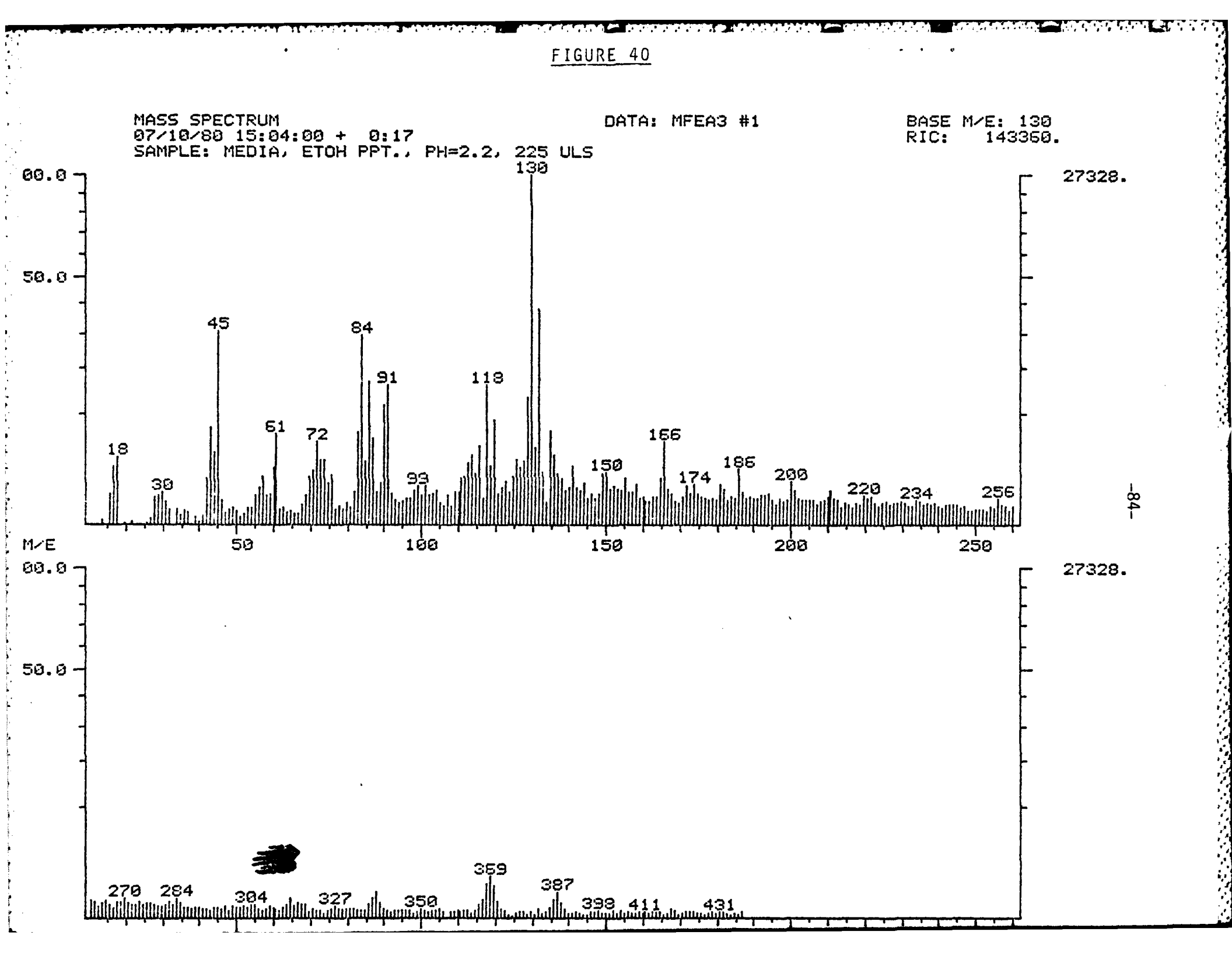
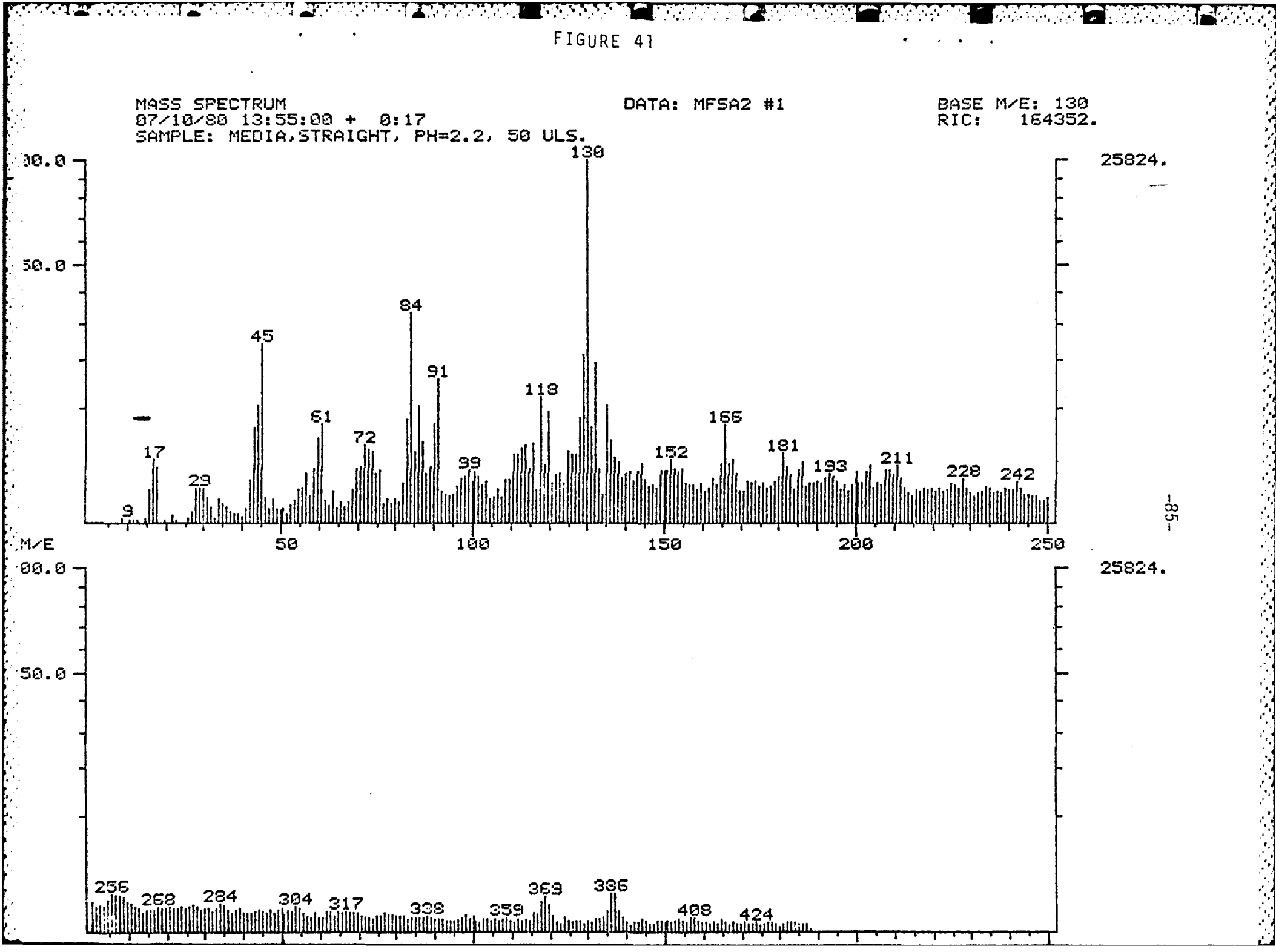

MASS SPECTRUM
07/10/80 15:04:00 + 0:17
SAMPLE: MEDIA, ETOH PPT., PH=2.2, 225 ULS
DATA: MFEA3 #1
BASE M/E: 130
RIC: 143360.
00.0
50.0
27328.
130
45
84
91
118
61
72
18
30
93
150
166
174
186
200
220
234
256
M/E
50
100
150
200
250
00.0
50.0
27328.
359
387
270
284
304
327
350
398
411
431

FIGURE 41


MASS SPECTRUM
07/10/80 13:55:00 + 0:17
SAMPLE: MEDIA,STRAIGHT, PH=2.2, 50 ULS.
DATA: MFSA2 #1
BASE M/E: 130
RIC: 164352.
100.0
50.0
25824.
130
84
45
91
118
61
72
166
152
181
193
211
228
242
17
29
99
9
M/E
50
100
150
200
250
100.0
50.0
25824.
256
268
284
304
317
338
359
369
386
408
424

The one ml thawed samples were treated with $HgCl_2$ to achieve a final concentration of 5 mM, which is sufficient to inactivate any virus particles. The pH was adjusted to 2.2 using NaOH or HCl. Two separate 50 microliter aliquots were taken and applied to strips of glass fiber filter paper. The samples were dried and stored at room temperature for periods of one day or less prior to analysis. Both field ionization and isobutane chemical ionization molecular profiles were obtained from this sample set.

In the case of the field ionization data, the zero hour samples were compared in all possible pairs for the control, low and high virus levels using the t-test for variable selection. Using a p-value threshold of 0.01 these tests gave between 6 and 16 m/e's out of 300 indicating a slight but significant difference between the zero hour patterns. This difference does not appear to be due to a photochemical effect of the UV inactivated media used in the preincubation of two of the cultures. A photochemical effect should be most apparent when comparing the control (100 percent UV treated0 and high virus (0 percent UV) samples. A similar pattern difference should be apparent when comparing the low virus (90 percent UV) to the high virus (0 percent UV). The first of the above t-tests gave 16 m/e's, the second pair gave only 6, and the third pair (control-100 percent UV vs low virus - 90 percent UV) gave 13 m/e's with $p \leq 0.01$ in each test. The small number of masses found in the second t-test and the fact that there was only 1 mass in common with the first test is inconsistent with a photochemical artifact. The three t-tests described gave 28 different m/e values with only 7 m/e's in common among the various pairs. Further studies will attempt to locate possible systematic errors in the experimental procedure that produce the observed zero hour differences. The zero hour differences that were observed in this experiment do not, however, have any common features with the more important metabolic or virus related pattern differences studied at other

incubation times. The possibility that there is an actual biological virus effect expressed during the pre-incubation period cannot be excluded at this point and it will be carefully examined in future experiments.

To select pattern changes due to the normal cell growth, a t-test comparing control samples at zero and 48 hours was used. Using 25 m/e's eight different discriminant analyses were performed. Two representative classifications are shown in Figures 42 and 43. First the time sequence samples were classified for each of the three virus levels 0, 10^3 or 10^4 TCID using the zero and 48 hour samples to derive the classification function. These results are shown in Figure 44. The figure gives the coordinates of the group averages for the most significant canonical variable derived from the classification function. The control samples show the largest range of values, consistent with the fact that the zero and 48 hour control sets were used for the initial variable selection.

The classification of the control group (A) thus reflects the progressive appearance of metabolic changes in the growth medium as a function of incubation time. The low virus samples (C) show a similar trend, although with a greatly reduced range of coordinate values and an apparent sequence reversal of the 6 and 12 hour samples. The high virus classification shows a slightly distorted sequence and an even smaller range of values. The last two classifications also show increasing overlap of samples between groups. This result is consistent with our previous experiments and shows a progressive pattern change due to cell growth that is inhibited by virus infection. The above sequence of classification also shows that the inhibition increases with the dose level of the virus.

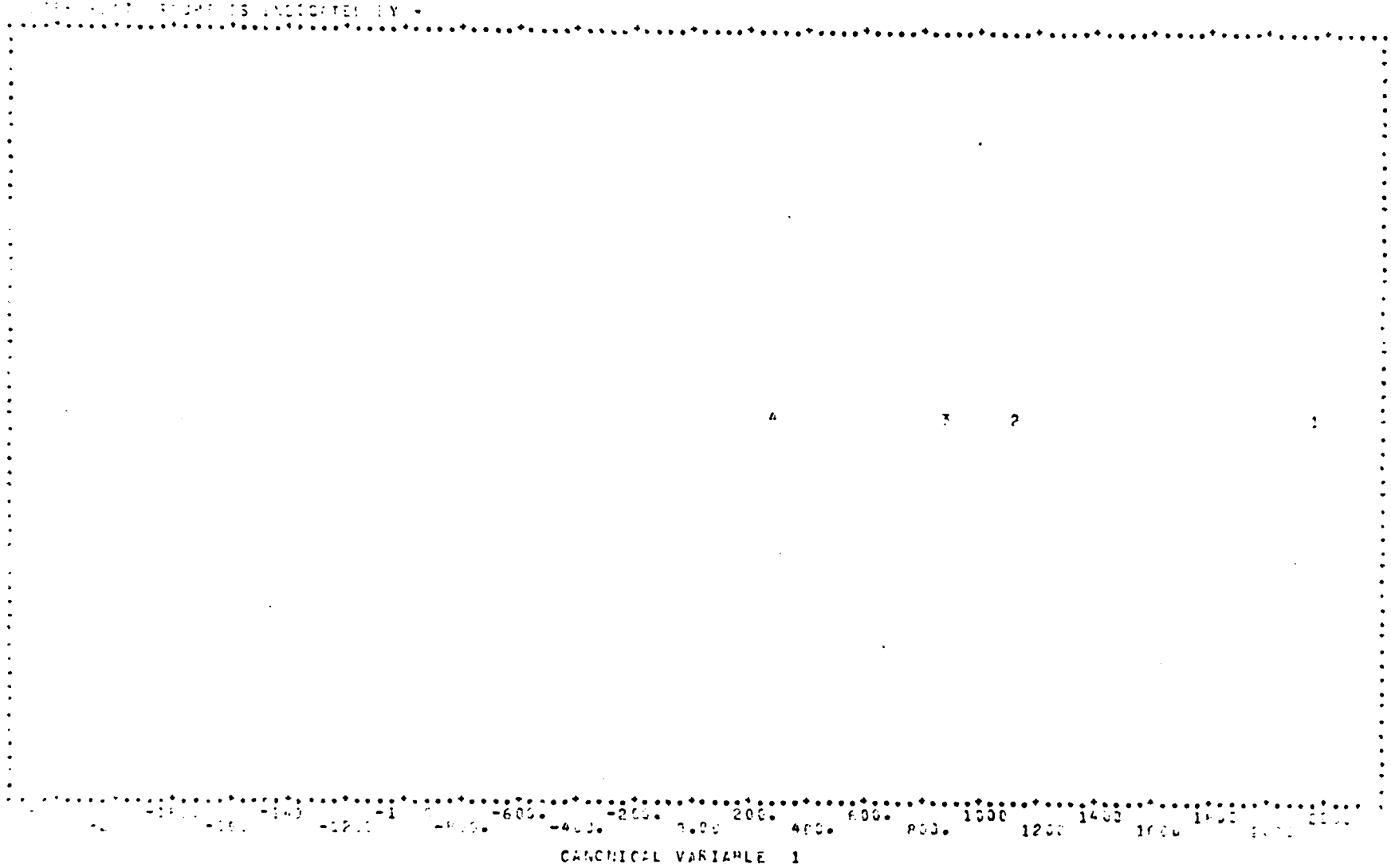


FIG. 42: Discriminant analysis classification of control tissue cultures versus incubation time. Numerals indicate group average locations: 1 = zero hr; 2 = 6 hr; 3 = 12 hr; 4 = 24 hr and 5 = 48 hr samples. m/e values selected from 48 hr change in control samples.

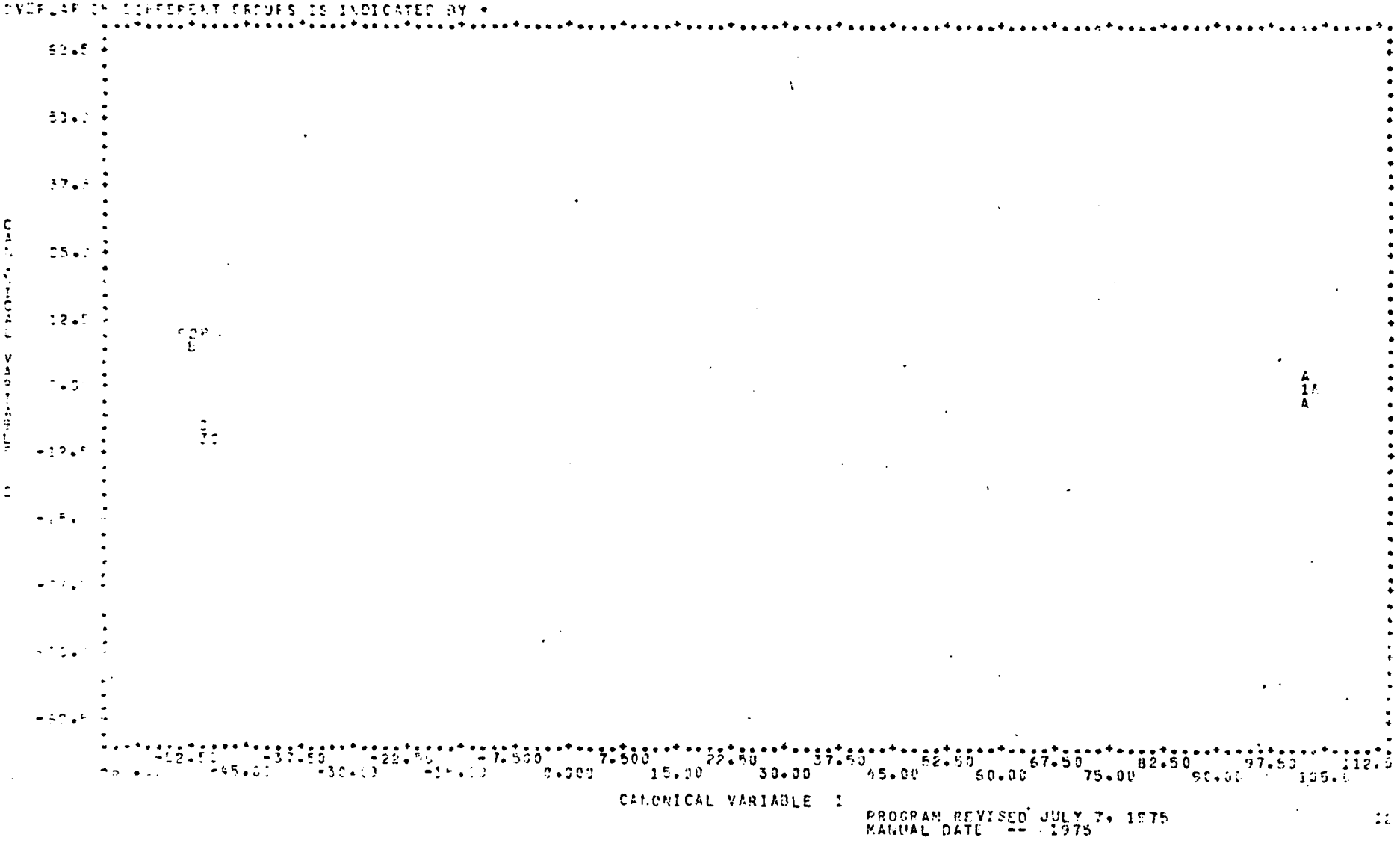


FIG. 43: Discriminant analysis classification of 0, 10^3 and 10^4 TCID virus level samples at six hours. Numerals indicate group averages and letters indicate individual sample locations: 1,A = 0 TCID (control); 2,B = 10^3 TCID (low virus); 3,C = 10^4 TCID (high virus). m/e values selected for 48 hr change in control samples.

FIGURE 44

Classification coordinates of group averages for discriminant analyses of experimental groups vs time. (The first row is the same classification shown in Figure 42). m/e values selected from 48 hr change in control samples.

Incubation Time (hrs)	0	6	12	24	48
Control Group, 0 TCID, A	2147	1133	872	260	-2000
Low Virus; 10^3 TCID, C	159	72	144	-25	-136
High Virus; 10^4 TCID, B	11	-23	0.13	-10	- 12

Next, five discriminant analyses were performed to classify the three virus levels at each individual incubation time. These results are shown in Figure 45. The table gives the distances between the group averages and the average or worst case spread out for the groups. The worst case value is given for groups having much larger variance with the subscript denoting the group with the largest spread. In this analysis the zero hour separations are smaller than those obtained for any other incubation time indicating again that the major experimental artifacts encountered in our first experiment have been largely removed by using UV inactivated cultures as the dilutant for all samples. Since the variables initially selected for this classification reflect primarily the normal metabolic processes the separation observed can be attributed to a differential inhibtition of these processes by the two virus levels. The first row in the table suggests that for the highest virus level the maximum pattern difference occurs at 12 hours. With the lower virus level the maximum signal is delayed to 24 hours. The third row showing the difference between the two virus levels is also maximized at 12 hours. In interpreting the data in this table it is important to note that each column represents a separate discriminant analysis so that distances should not be compared along a row without also relating them to the other figures in their respective columns.

Another set of 25 variables were selected using the t-test to compare the samples infected with the highest virus level at zero and 48 hours. Using these variables, discriminant analyses were performed as above to classify the three different time sequences and to classify the three virus levels at each incubation time.

FIGURE 45

Distance between group averages from discriminant analysis classification of different virus levels at various incubation times. m/e values selected from 48 hours changes in control group (A).

	Incubation Time (hrs)				
Group[a]	0	6[c]	12	24	48
A-B	11	152	1140	630	90
A-C	23	150	180*	567	45*
B-C	25	25	1320	66	44*
S[b]	3	2	160	2	70_C

a) Group designations: A = control; 0 TCID virus level; B = high 10^4TCID virus level; C = low, 10^4 TCID virus level.

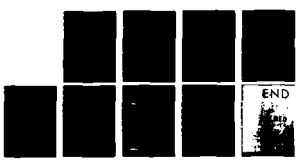
b) Average within group spread; subscript indicates worst case single group spread.

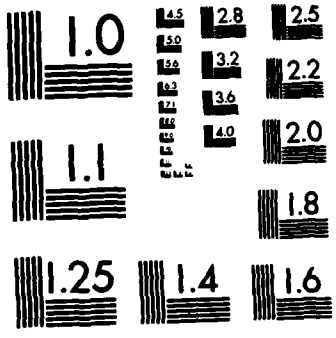
c) This row is the same classification shown in Figure 43.

* Indicates incomplete separation between groups.

AD-A126 763 MASS SPECTROMETRIC RAPID DIAGNOSIS OF INFECTIOUS DISEASES(U) STATE UNIV OF NEW YORK AT BUFFALO SCHOOL OF MEDICINE M ANBAR ET AL. MAR 81 DAMD17-78-C-8035 2/2

UNCLASSIFIED F/G 6/5 NL





There was only one m/e vaue in common between peaks selected by this t test and by previous t test. This drastic change may indicate either that the virus infected cells have a significantly altered metabolic profile or that the normal cells have an altered metabolism at 48 hours compared with ealier times due to self inhibition from excreted products or depletion of nutrients.

For this set of masses the time sequence classifications became more confused, with considerable overlap between groups throughout the incubation series. This is not surprising in view of the lack of overlap between this set of variables and the previous one, presumably reflecting normal metabolic changes.

The classification of different virus levels at each time is considerably improved using this new set of variables as seen in Figure 46. The first feature to note is that the variance within each group is consistently small with the exception of the zero hour classification where the results indicate the expected overlap of all three groups. As before the separation between different virus levels is strongest at earlier times and is maximized at 6 to 12 hours.

This analysis was extended to include t test of seven additional group sets, using different times of incubation for zero and high virus levels. Figure 47 summarizes the results of discriminant analysis separation of different virus levels at each incubation time using variables from eight different t-tests. All t-tests yield variables able to separate samples from a given time according to virus level. Variables selected at six hours generally do not lead to as large a separation as other time points.

Using each of these different sets of variables a significant group separation was found in each case between the "zero" time sets at different virus levels. These results indicate a real difference between these sets, suggesting that measurable metabolic differences occur already during the one

FIGURE 46

Distances between group averages from discriminant analysis classification of different virus levels at various incubation times. m/e values selected from 48 hour changes in high virus group (P).

Incubation time (hrs)	0	6	12	24	48
Groups[a]					
A-B	100	265	177	12	56
A-C	59*	103	52	31	52
B-C	41*	162	228	20	77
S[b]	110_C	3	3	3	4

a) Group designations: A = control; OTCID virus level; B = high 10^4 TCID virus level; C = low, 10^4 TCID virus level.

b) Average within group spread; subscript indicates worst case single group spread.

* Indicates incomplete separation between groups.

FIGURE 47

Summary of separation data obtained from eight different variable sets used to separate three virus levels at each incubation time. Table entries are computed from distances between group averages obtained by discriminant analysis.

$$\text{(Separation entry} = \log \left[\frac{\text{sum of between group distances}}{\text{average group spread}}\right])$$

Variable Selection	Incubation Time (hours)				
	0	6	12	24	48
Control; OHR vs: 48 hr	1.3	2.2	1.2*	2.8	.4*
24 hr	.9	2.8	2.0	2.2	2.3
12 hr	2.6	2.3	.8*	2.9	2.4
6 hr	2.4	1.4	.8	1.3	.6*
High virus; OHR vs: 48 hr	.2*	2.2	2.2	1.3	1.6
24 hr	2.3	3.0	2.6	.8*	1.2*
12 hr	1.9	2.2	2.0	.8*	3.0
6 hr	2.0	2.4	.6*	2.6	1.8*

* Indicates number of incomplete group separations in discriminant analyses.

hour "preincubation period".

An examination of the masses selected by the four control group t-tests showed only 3 masses in common, using the m/e's with the 25 lowest p-values from each test. For m/e's with $p \leq 0.05$ in all tests there were 13 common variables. This rather low number of common variables means that different metabolic processes differentiate best between infected and non-infected cells and between cells infected at different levels of viral infection, at different times of incubation. It is, therefore, difficult to distinguish between metabolites primarily associated with the viral infection, and those primarily characteristic of normal cell turnover, which is inhibited by the infection.

The low degree of overlap between the variables selected by various t-tests led us to apply an alternate variable selection for this type of multi-parameter experiment. The BMDP program P2V was used for analysis of variance of m/e intensity measurements, using time and virus level as the grouping factors and examining the individual m/e's as dependent variables. The virus level group consisted of high and control (non-viable) virus levels and the time group consisted of 0, 6, 12, 24, and 48 hour incubations. This program analyzes the dependent variables for effects due to either virus level, time, or the combination of both. Using this program 33 m/e's were selected with a level factor significant at $p \leq 0.05$. Figure 48 shows the separation between virus levels at different times using these variables for discriminant analysis. A similar degree of separation to that obtained with the t-test selection was obtained.

A more rigorous test of the use of these 33 variables was obtained by using them to derive a classification for both time and virus level simultaneously. Figure 49 shows the results of a discriminant analysis of zero, 6, and 12 hour samples of high and zero virus levels, demonstrating

FIGURE 48

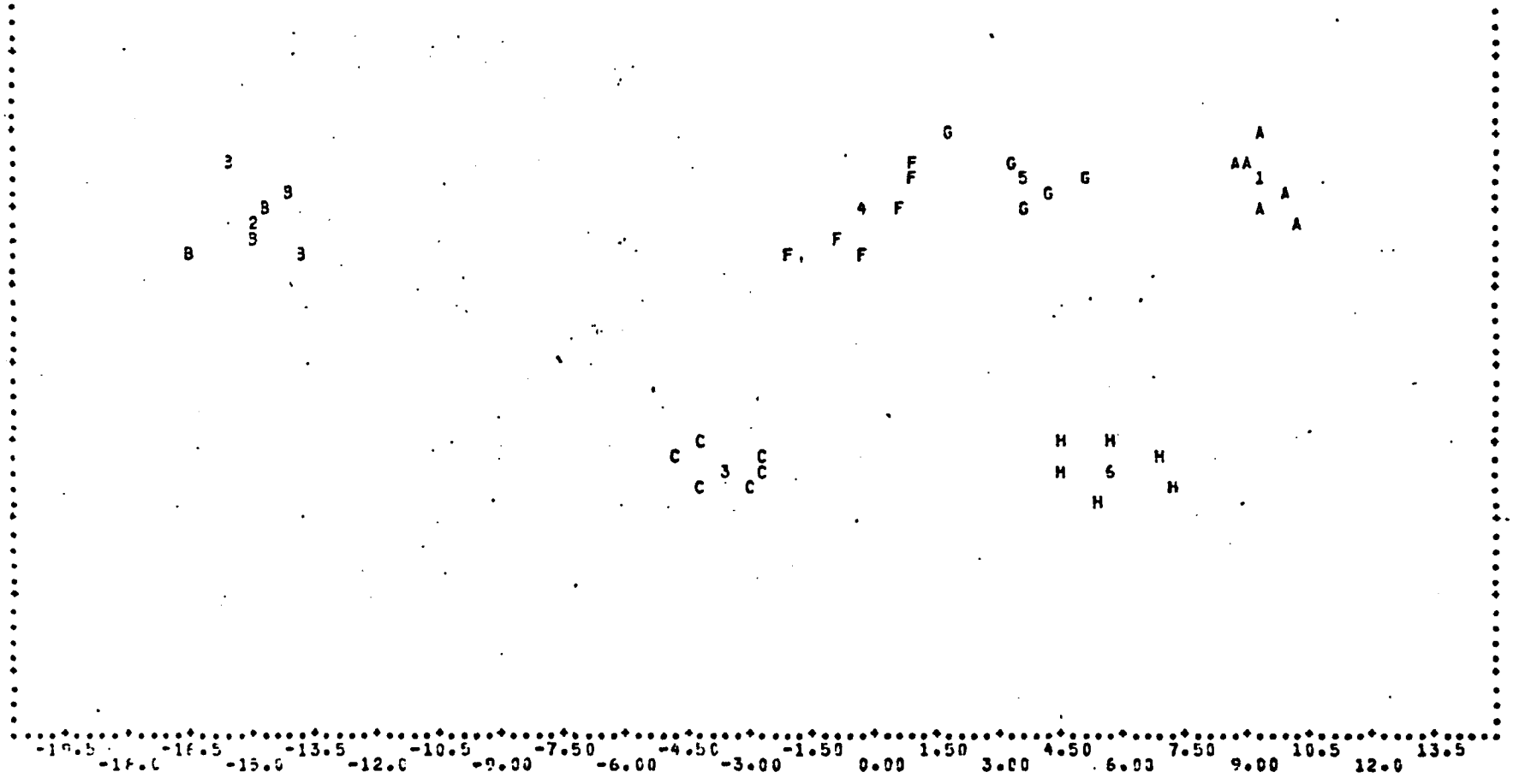
Distance between group averages for discriminant analyses of experimental groups vs time. Variables (m/e values) selected by analysis of variance with grouping factors time and virus levels.

	Incubation Time (hours)				
Group[a]	0	6	12	24	48
A-B	360	387	90	1460	55
A-C	285	495	112	1580	201
B-C	74	882	120	123	146
S[b]	3	2	3	3	3

a) Group designations: A = control, (0 TCID) virus level; B = high, (10^4 TCID) virus level; C = low (10^3 TCID) virus level.

b) Average within group spread.

FIG.49 Discriminant analysis of high and zero virus level samples from 0, 6, and 12 hour incubations using 32 variables from program P2V. Numerals denote group average locations.



A = 0HR control; B = 6HR control; C = 12 hr control

F = 0HR high virus; G = 6HR high input; H = 12 HR high virus.

complete separation of all six groups. Whereas the infected samples at 6, 12, and 24 hours still separated well, there was a tendency of all three zero hour groups to form a cluster which partly overlapped the clusters of the infected groups at 6, 12, and 24 hours. Also the separation of the 48 hour infected samples from the other groups was smaller than that of samples with shorter times of incubation.

These findings corroborate our previous suggestion that significant metabolic changes due to viral infection of tissue cultured cells can be monitored in the first 12 hours and possibly as soon as in the first hour or two. Furthermore, there seems to be no gain in diagnostic information by incubation for 24 or 48 hours. Also it seems that lower levels of infection than those used in these preliminary experiments could be detected within the same time frame.

The chemical ionization patterns for zero and 48 hour cultures were analyzed using the t-test and discriminant analysis with similar results as obtained from the FI data for the same samples. Other CI samples in this experimental series (6, 12, 24 hour samples, analyzed on different days) showed a drastic drop in the source sensitivity, making these patterns totally different from the other samples. Additional experiments will be undertaken to establish the proper control of CI parameters for reproducible profile analyses.

In the latest study, Mahoney strain of Polio virus was cultured in BGM (Buffalo Green Monkey) cells as follows: a stock solution of the virus was diluted a thousand fold. One ml was added to a 75 cm^2 flask containing a cofluent monolayer of BGM cells. The virus was absorbed for an hour. The cells were fed with 10 ml of Eagle's Minimum Essential Medium (MEM) supplemented in 2 percent Bovine calf serum and pH adjusted using sodium bicarbonate.

The cells were allowed to incubate at 37°C for 48 hours, until microscopic examination showed that 90 percent or more of the monolayer appeared to be infected. The cells remaining attached to the flask were scraped off and along with the media were transferred to a tube and centrifuged at 2000 rpm for ten minutes at 4°C. The supernate was collected and divided into 2 ml aliquots and snap frozen at -70°C.

One aliquot was used to titer the virus. This was done by serially diluting the virus pool ten-fold until 10^{-10} dilution of the original pool was achieved using 1.8 ml of MEM and 0.2 ml of virus. Each dilution was innoculated in triplicate using 0.1 ml of virus per tube of GM cells. The tubes were innoculated as the flask was and were observed over a 72 hour period. At that time the tissue culture infective dose (TCID50) was determined to be $10^{7.5}$ per ml.

Two 2 ml aliquots of the virus pool were placed in two shallow petri dishes and exposed to U.V. light for a time of 10 and 30 minutes, respectively, at a distance of 12.5 cm. The virus was then innoculated into the BGM cells and observed over a 72 hour period for cytopathic effect (CPE). No CPE was observed in either the 10 or 30 minute innoculated tubes. Ten minutes of U.V. irradiation appeared to be a sufficient exposure time to inactivate the polio virus.

Two ml of the polio virus pool were diluted by one-tenth with MEM. Three ml of this were set aside and the remaining 17 ml were U.V. irradiated as described above. Three serial ten fold dilutions of the tenth diluted virus pool were done using the irradiated pool as the diluent. This provided three virus pools: (1) U.V. inactivated, (2) high titer (approximately $10^{6.5}TID_{50}$, with a multiplicity of infection (MOI) of 10), (3) low titer (approximately $10^{3.5}TID_{50}$, with a MOI of 0.1). There were approximately 3.5×10^5 cells/tube. The growth media was removed from 45 tubes and

replaced with 0.1 ml of virus (that is, 15 tubes per virus pool).

Three tubes from each virus pool were fed 1 ml of MEM and immediately centrifuged at 2000 rpm for 5 minutes and the supernates were collected. These were considered "time zero". The rest of the tubes were placed in an incubator at $37^{o}C$ for 1 hour to allow all .1 ml of virus to adsorb. These tubes were then fed 1 ml of MEM. Three tubes from each virus pool were immediately centrifuged and the supernates were collected and labelled T1. The rest of the tubes were incubated at $37^{o}C$ centrifuged, and the supernates collected 2, 4 and 6 hours after innoculation. Tubes from the 4 and 6 hours were observed for any CPE. After 4 hours no CPE was observed in either the U.V. inactivated or the low titer tubes. However, in the high titer tubes less than 10 percent of the cells showed rounding or swelling which might be described as CPE due to the polio virus. After 6 hours there were no changes seen in any of the tubes.

The tubes were frozen at $-80^{o}C$ until analysis. The one ml thawed samples were tested with 20 microliters of 0.25 M $HgCL_2$ to achieve a final concentration of 5 mM, which is sufficient to inactivate any virus particles. The pH was adjusted to 2.2 using NaOH or HCl. Two separate 40 microliter aliquots were taken and applied to strips of glass fiber filter paper. The samples were dried and stored at room temperature for periods of one day or less prior to analysis.

As in the previous study, there were triplicate tubes that were run in duplicate, thereby giving six samples for each virus concentration and each time period. Each of these individual groups of six samples were analyzed for their respective reproducibility. In this study, the average coefficient of variation for each of the individual groups was approximately 10 percent worse than in the previous study (25 percent vs. 15 percent over the mass ranger of 61 to 361 amu; 16 percent vs. 11 percent over the mass range of 100 to 250

amu.). An effort is now underway to determine the reason for this difference.

One of the most striking aspects of this new study was the difference in the number of masses with $p < 0.05$ in the control group using the t-test of zero hour versus six hours (see Table 1). In the previous study there were 81 masses with $p < 0.05$ in a mass range of 61 to 361 compared with just 21 masses in the latest study. In the case of the high virus levels, the number of masses were similar in both cases, but again in the latest effort there were fewer masses (26 vs. 29). Another disturbing feature was the fact that there was still a "zero hour effect", but it was smaller this time.

Many parameters have changed since the previous study so it is difficult to compare experiements. It was indeed unfortunate that the original cell line (human embryonic lung tissue) died and BGM (Buffalo Green Monkey) cells were substituted. Certainly, the type and rate of metabolic changes would be different for various cell lines. This could partially explain why there was such a difference between the control samples of each study. Clearly, there was less activity in the control group this time - possibly attributable to the slower rate of growth of the BGM cell (six hours is not enough time) or the UV inactivated virus did effect cell metabolism. The higher coefficient of variation would undoubtedly have decreased the number of masses with low p-values.

Overall, the results of the latest study were somewhat disappointing. The trends seen in the earlier studies were less distinctive in this recent effort. It is obvious that the higher coefficient of variation contributed to this result. Analysis of the pooled freeze-dried urine that is used by this laboratory as a monitor of pattern reproducibility indicated that the average coefficient of variation due to the mass spectrometer had not changed from the previous studies. Therefore, it would seem that the tissue culture experimental techniques have to be improved. A possible way to monitor this

FIGURE 50

Number of $p < 0.05$

For	Mass Range			
	61–100	100–200	200–300	300–360
Control (U.V. inactivated virus) 0 hrs vs 6 hrs, previous study	10	37	29	5
Control (U.V. inactivated virus) 0 hr vs 6 hrs, this study	4	9	6	2
High Virus ($10^4 TCID_{50}$) 0 hr vs 6 hrs, previous study	7	13	7	2
High Virus ($10^{6.5} TCID_{50}$) 0 hrs vs 6 hrs, this study	3	10	9	4

methodology would be to make each cell culture as its own control. Instead of 45 tubes containing cell cultures, 9 tubes would be used - three with UV inactivated virus, three with low titer virus, and three with high virus. All tubes would be sampled immediately after the addition of virus and then incubated. Aliquots of the defined media from each tube would be taken at designated times. A careful study of each tube would be recorded using the standard tissue culture techniques - that is the cell growth for the UV inactivated virus cultures would be measured and the CPE effects for the virus infected cultures would be noted. With such short time periods, it obviously would be important to verify that indeed cell growth had taken place in the uninfected cultures and that the virus had infected the infected cells. All of our virus studies to date have indicated that virus infection can be detected in vitro within a few hours after exposure. This is consistent with tracer and other virological experiments which have shown that cells innoculated with poliovirus experienced pathological changes within hours after infection.

[Reprinted from CLINICAL CHEMISTRY, 26, 1443 (1980).]

Diagnosis of Viral Infections by Multicomponent Mass Spectrometric Analysis

R. Abbott, M. Anbar, H. Faden, J. McReynolds, W. Rieth, M. Scanlon, L. Verkh, and B. Wolff

Metabolic profiles of urine extracts of humans with viral infections, as well as of media of virus-infected human tissue cultures, have been analyzed by non-fragmenting mass spectrometry and compared with corresponding controls. The spectra were then subjected to several alternative computerized statistical procedures to detect diagnostic biochemical profiles. Controlled longitudinal studies on fully informed, consenting volunteers who received sandfly fever virus demonstrate the onset of a characteristic metabolic pattern that precedes the onset of symptoms and subsides when the patients overcome the infection. Longitudinal studies of human tissue cultures infected with poliomyelitis virus demonstrate characteristic metabolic patterns within a few hours after infection. Non-fragmenting mass spectrometry may thus provide the clinical laboratory with a sensitive, reliable test for viral infections significantly faster than attainable by current techniques.

Additional Keyphrases: *metabolic profiling · urine · viruses · sandfly fever · poliomyelitis · data processing*

Diagnostic metabolic profiling of biological fluids as described by other workers using a variety of techniques (*1–5*) can also be obtained by non-fragmenting mass spectrometry (*6*). The advantage of non-fragmenting mass spectrometry is its ability to quantitate hundreds of constituents in a biological sample within a relatively short time (less than 1 h), expressing abundances in a digital form ready for on-line computer analysis. No other analytical technique matches these capabilities. Multicomponent metabolic profiles lend themselves to statistical pattern-recognition analysis in which a characteristic pattern of a large number of constituents is used to classify the biological samples and thereby to diagnose pathological states.

A previous report (*6*) described the use of field-ionization mass spectrometry for the multicomponent analysis of urine and the identification of patients with an acute liver disorder (infectious hepatitis). The metabolic aberrations associated with liver dysfunction are extensive, and the characteristic changes in urine composition that occur could be recognized by several alternative clinical chemistry techniques. The potential diagnostic usefulness of mass-spectrometric metabolic profile analysis thus awaited a more challenging problem.

We have since improved our field-ionization mass-spectrometric instrumentation (*7*), and have compared it with chemical-ionization mass spectrometry. We improved considerably the sample preseparation treatment, and have also explored the possibility of minimizing these procedures, limiting them to the enzymic removal of urea. We thus compared the diagnostic information obtainable from the major constituents in urine with that obtained from a certain subset of separated metabolites, which includes many minor constituents. We have also tested and compared alternative statistical data-handling procedures for the diagnostic classification of metabolic profiles. The previously reported method, a Wilcoxon test followed by computation of a weighted noncorrelation index for comparison of patterns, has been replaced with a set of programs from the BMPD package (*8*), which have greater flexibility for the treatment of multiple group classifications. These improvements in methodology have resulted in a more sensitive and specific diagnostic technique, enabling us to identify characteristic metabolic profiles in urines of children with pneumonia or with virus-induced diarrhea.

A more critical test of the technique, however, was the identification of transient metabolic changes induced in human subjects after inoculation with sandfly fever virus to produce a mild, self-limited infection. These transient changes were identified in a longitudinal study over a period of one month, each subject serving as his own control. In this paper we also describe preliminary results of the in vitro detection of the presence of a virus in a biological sample by characteristic metabolic profile changes in the medium of an infected tissue culture. Our findings thus corroborate our preliminary suggestion (*6*) that non-fragmenting mass spectrometry has the potential of becoming a highly powerful diagnostic tool in the clinical laboratory.

Materials and Methods

Longitudinal Study of Virus Infection

Nine fully informed, consenting volunteers participated in this study, which was conducted by the U.S. Army Medical Research Institute for Infectious Disease (USAMRIID). During the study subjects stayed in a hospital ward and received a similar diet. Seven individuals received sandfly fever virus and two individuals received placebo injections. The participants were not told whether they received the virus or a control injection. Morning urine samples were collected four days before the injection, the day of the injection, for eight consecutive days thereafter, and finally 28 days after the injection. Samples were frozen without preservative, shipped to our laboratory packed in solid CO_2, and kept frozen until analysis. The study with volunteers was conducted as a portion of long-term investigations at USAMRIID concerned with the diagnosis, prevention, and treatment of infectious diseases, and was extensively reviewed and approved in accordance with existing U.S. Army Regulations. The collection of urine samples for various clinical tests allowed us to acquire the specimens described herein.

Tissue-Culture Virus Infection Study

A cell line of human embryonic lung tissue was cultured in Eagle's minimal essential growth medium supplemented with 50 mL of newborn calf serum per liter. Cultures with equal numbers of cells, prepared in 5-mL plastic culture tubes, were incubated at 37 °C. Three tubes each were prepared for analysis at 0, 6, and 24 h for both infected and uninfected cultures. Virus-infected cultures were prepared by adding 0.1 mL of medium containing a tissue-culture infective dose

Department of Biophysical Sciences, School of Medicine, State University of New York at Buffalo, Buffalo, NY 14214.
Received Apr. 24, 1980; accepted June 26, 1980.

Table 1. Stepwise Discriminant Analysis of Longitudinal Samples from Two Individuals Receiving Sandfly Fever Virus, Based on 40 *m/z* Values (Variables)

Sample	Day	Indiv.	Class [a]	Mahalanobis D^2, and posterior probability for group [b] Control	Fever
Group A					
1–8	−4, 0	J + L		14.0, 1	—, 0
Group C					
13–22	3, 4, 5	J + L		—, 0	14.4, 1
Group B					
9	2	J	F	—, 0	—, 1
10	2	J	F	—, 0	—, 1
11	2	L	C	—, 1	—, 0
12	2	L	C	—, 1	—, 0
Group D					
23	7	J	C	—, 1	—, 0
26	7	J	F	—, 0	—, 1
24	8	J	C	—, 1	—, 0
27	8	J	C	—, 1	—, 0
25	28	J	C	—, 1	—, 0
28	28	J	C	—, 1	—, 0
29	7	L	F	—, 0	—, 1
32	7	L	F	—, 0	—, 1
30	8	L	F	—, 0	—, 1
33	8	L	F	—, 0	—, 1
31	28	L	C	—, 0	—, 0
34	28	L	C	—, 1	—, 0

[a] Classification function based on days −4, 0 (Control) and 3, 4, 5 (Fever) used to classify samples from all other days.
[b] — indicates D^2 greater than 10^4.

range 1 to 450 were acquired and stored by the data system. The probe and source were baked to 250 °C for 10 min between each sample. After bakeout, the source temperature was quickly returned to 200 °C by using a copper-tipped solid probe to conduct heat away, through a heat pipe. Because the source high voltage was turned off during cooling, and because the calibration mix itself "washes" adsorbed components out of the source, a new calibration was obtained for each sample.

Chemical-Ionization Mass Spectrometry

Chemical-ionization mass spectra were obtained with a DuPont 21-491B mass spectrometer (DuPont Instruments, Wilmington, DE 19898) equipped with a chemical-ionization source; isobutane reagent gas was supplied at a source pressure of 0.3 to 0.5 Torr (40–70 Pa). Mass resolution of 500 and a 10-s cyclic scan from 1 to 550 amu gave integrated counts of 1.5×10^5 to 2×10^5 for 1 μg of adenine evaporated from the probe, corresponding to 30 pC/μg sensitivity. For analysis of urease-digested urine samples we programmed the solid probe input power to evaporate the entire sample within 45 scans in 7.5 min. The Finnigan-Incos Model 2400 data system was used also with this instrument to acquire and store spectra.

Data Processing

Raw spectra, consisting of m/z centroids and peak areas acquired in single scans, were added together by the data system to give an integrated spectrum of the total number of ions detected at each nominal mass during the evaporation of the sample. Added spectra were transmitted to a CYBER 173 central computer for further processing (Control Data Corp., Minneapolis, MN 55440). Each spectrum was normalized to unit area by excluding from the normalization sum all peaks greater than 5% of the total ion count (*6*). We used the normalized spectra for statistical analysis.

The M_r profiles of the different sample groups previously described have been analyzed by several multivariate statistical programs of the BMDP package (*8*). Programs P7D (description of groups with analysis of variance), P3D (t-test), P2M (clustering analysis), and P7M (stepwise discriminant analysis) were used for different steps of the statistical analysis of the diagnostic patterns. These programs were documented in a manual published by UCLA, and only a brief description of the function of each program will be given here.

The P7D program is used to display the mean and variation of individual mass intensities for two or more persons as a function of the sampling day in the longitudinal study. The P3D program tests the null hypothesis that each of the variables in two tested groups belongs to the same population. The clustering analysis program P2M calculates the n-space euclidian distance between each spectrum (case), by using the scaled normalized intensity of chosen m/z values (variables) as single coordinates. From this distance matrix the program constructs a diagram showing the relative proximity of samples in the space determined by the selected variables.

The P7M stepwise discriminant analysis program constructs an optimized classification function as a linear combination of variables that achieves the best separation of the given spectra into a specified number of groups. Specifically, the program maximizes the ratio of sum of squares of the between-group variance to that of the within-group variance. As each variable is added into the classification function, the program recalculates an F-statistic for all variables, adding variables that exceed a specific "F to enter" or removing variables that fall below a specified "F to remove." The multiple correlation coefficient of entered variables with previ-

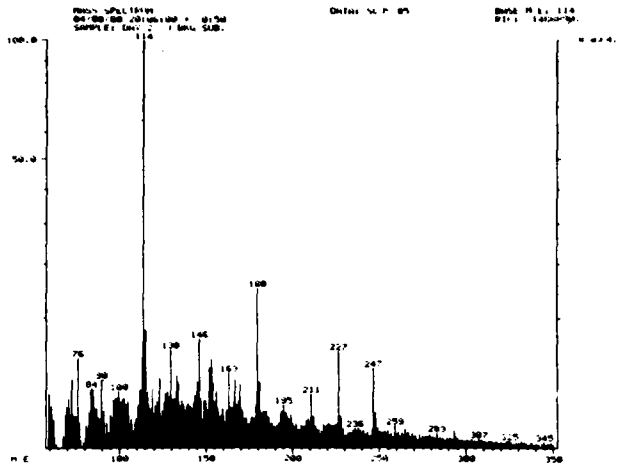


Fig. 2. Isobutane chemical-ionization integrated mass spectrum of 1 μL of urease treated urine

by day 7. The observed pattern differences in morning urine samples thus preceded the appearance of clinical symptoms and persisted after their disappearance. The third set of samples, from a control subject, was classified as healthy throughout by the same classification procedure. Samples from the other six participants in this study were recently obtained from USAMRIID, and work is in progress to obtain the field-ionization profiles of extracted metabolites from these samples.

Because this method requires considerable care and effort, we are limited to the analysis of 40 to 50 samples per week; consequently, we also tested a simplified profiling procedure consisting of analysis of the urease-digested urine by isobutane chemical-ionization mass spectrometry. The sample size (1 μL) was empirically established as that which would allow a relatively rapid sample evaporation while still maintaining a large excess in the isobutane reagent ion intensity. The longitudinal sample sets for each individual were prepared as described above and analyzed in duplicate in a random sequence. For each sample set an aliquot of the corresponding urine from a healthy subject was incubated with enzyme at the same time. A sample of this standard urine and a water blank containing the enzyme and buffer were analyzed at the beginning of each sample set, and a replicate of the standard sample was repeated at the completion of each set of subject samples.

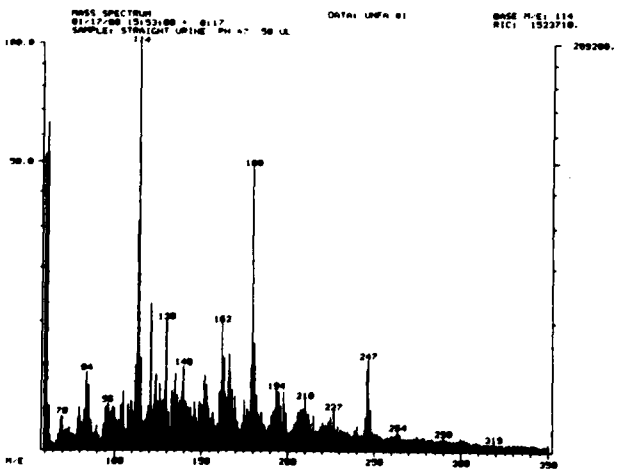


Fig. 3. Field-ionization integrated mass spectrum of 50 μL of untreated urine

Certainly, we achieved one objective of this method—a high sample throughput—in that the assay of the entire set of longitudinal samples from all nine participants, approximately 200 samples including blanks and standards, was completed in one week.

The patterns obtained were less informative than the data from the extracted samples. This is not surprising, because the sample size used limits the constituents examined to species excreted at more than 1 mg per day. A representative spectrum (Figure 2) is dominated by the protonated molecular ions of creatinine (m/z 114) and hippuric acid (m/z 180). Direct examination of urine by field ionization of 50–75 μL samples (Figure 3) gives an almost identical pattern except for the appearance of some constituents as non-protonated molecular ions (1 amu lower).

Analyses of the chemical-ionization mass-spectrometric patterns by the same programs applied to the field-ionization data indicated that all samples are rather similar; no unambiguous differentiation of the "response" period was obtained. A weak separation was, in fact, obtained with the chemical-ionization patterns. A t-test selection of variables and discriminant analyses of a subset of single replicate samples from all infected subjects resulted in a classification function that separated the control days (−4 and 0) from the fever days (4 and 5). Classification of each person's sample series by this classification gave 8% false positives and 27% false negatives. This result reflects the low signal-to-noise ratio of diagnostic information in this completely nonselective pattern of major urinary constituents, a ratio that is not expected to be significantly affected by the mild infection resulting from the experimental virus illness.

Although these findings demonstrate the inadequacy of the simplified sample preparation procedure for the detection of subtle metabolic changes associated with the infection, they indicate that, at least in terms of the major urinary constituents, no major dietary or environmental differences have been represented in this sample set. Thus the more subtle pattern differences detected in the extracted samples are obtained in the context of a normal, unperturbed baseline pattern.

The analytical and statistical procedures reported here have considerable potential for clinical diagnosis. In particular the high analytical throughput need not be sacrificed when more selective sample preparation methods are utilized to increase the diagnostic information. A certain amount of chemical selectivity without elaborate preparation could also be achieved by the use of other reagent gases (*10*).

The statistical procedures we used might be needed only for establishing the initial diagnostic criteria for a particular disease. The classification function could then be utilized in a minicomputer data-acquisition system, to rapidly classify unknown samples.

Virus-Infected Tissue Cultures

M_r profiles of media from cultures incubated for 24 h were prepared at pH 2, 7, and 10 to determine the optimum conditions for classification. Masses selected by the t-test procedure were used as variables for the P7M discriminant analysis program. We classified 0- and 24-h samples of infected and noninfected cultures at different pH values into four groups, to select the pH range giving the best separation; pH 2 was the most effective. Analyses of infected and uninfected samples from 6-h incubation at pH 2 are shown in Figure 4.

The initial classification revealed three major trends in the patterns. First, small differences in the 0-h patterns suggest a memory effect, reflecting composition differences in the media applied during the virus-absorption period. Because the intermediate wash is mild enough not to remove adsorbed virus particles, low M_r metabolic constituents from the in-

Two additional t-tests were used to obtain new groups of m/z values. By performing the t-test on 0 vs 24 h, we obtained a set of peaks that reflected both time- and virus-dependent pattern changes. The initial set of 42 variables was finally reduced to 25 by deleting 17 m/z values that showed significant differences in the 0-h t-test. For these 25 variables the discriminant analysis program derived a classification function to separate samples into five groups: A, all 0-h samples; B, 6-h infected samples; C, 24-h infected samples; Y, 6-h noninfected samples; and Z, 24-h noninfected samples (Figure 5). The probability of assignment of individual spectra to the corresponding group is presented in Table 2.

The zero time difference observed in these experiments can be avoided by preincubating control samples with an ultraviolet-radiation-inactivated aliquot of the virus culture used for infected samples. A more difficult task will be the deconvolution of possible virus-specific changes in patterns from the unavoidable cell-growth-dependent changes. We intend to extend this feasibility study by further experiments at additional time points with different viruses and host cell lines as well as cultures with intermediate concentrations of viruses. These studies should clarify the potential value of M_r profile analysis for early virus detection in tissue cultures.

Since our last report we have demonstrated excellent results in the classification of additional liver disorders (cirrhosis, in particular), pneumonia, viral-induced diarrhea, and urinary infections (unpublished results); the biochemical aberrations in all these diseases were, however, rather conspicuous, and the mass-spectrometric multicomponent analysis did not offer a unique diagnostic solution. The results reported in this paper demonstrate a unique capability of our methodology to identify minute changes in metabolic profiles with a high degree of certainty. Moreover, these changes were demonstrable before the onset of clinical symptoms, and persisted for some time after the clinical symptoms have subsided. The diagnostic implications of these findings are self-evident. As indicated by preliminary results, the analysis of metabolic profiles of tissue-culture media as a means to detect and possibly identify a viral infection much earlier than feasible by the current techniques is another highly promising application of multicomponent analysis by non-fragmenting mass spectrometry.

This study was sponsored in part by the U.S. Army Medical Research and Development Command under Contract No. DAMD177808035 and a grant from the Whitaker Foundation. The BMDP programs were developed at the Health Sciences Computing Facility, UCLA, under NIH Special Research Resources Grant RR-3.

References

1. Horning, E. C., and Horning, M. G., Metabolic profiles: Gas-phase methods for analysis of metabolites. *Clin. Chem.* **17,** 802 (1971).

2. Jellum, E., Stokke, D., and Eldjarn, L., Combined use of gas chromatography, mass spectrometry and computer in diagnosis and studies of metabolic disorders. *Clin. Chem.* **18,** 800 (1972).

3. Zlatkis, A., Lee, K. Y., Poole, C. F., and Holzer, G., Capillary column gas chromatographic profile analysis of volatile compounds in sera of normal and virus-infected patients. *J. Chromatogr.* **163,** 125 (1979).

4. Gates, S. C., Smisko, M., Aschendel, C., et al., Automated simultaneous qualitative and quantitative analysis of complex organic mixtures with a gas chromatography–mass spectrometry–computer system. *Anal. Chem.* **50,** 433 (1978).

5. Issachar, D., and Yinon, J., Screening of organic acids in urine by chemical ionization mass spectrometry. *Biomed. Mass Spectrom.* **6,** 47 (1979).

6. Anbar, M., Dyer, R. L., and Scolnick, M. E., Diagnosis of infectious hepatitis by multicomponent analysis. *Clin. Chem.* **22,** 1503 (1976).

7. McReynolds, J. H., Flynn, N. W., Sperry, R. R., et al., Glass lined field ionization sources. *Anal. Chem.* **49,** 2121 (1977).

8. BMDP, Health Sciences Computing Facility, University of California at Los Angeles; program date July 7, 1975.

9. McReynolds, J. H., and Anbar, M., Combined field ionization ion kinetic energy spectrometer multicomponent analysis. *Int. J. Mass Spectrom. Ion Phys.* **24,** 37 (1977).

10. Hunt, D. F., Reagent gases for chemical ionization mass spectrometry. *Adv. Mass Spectrom.* **6,** 517 (1974).

END

FILMED

5-83

DTIC